



PHD

**Helical constraints and in silico library screening to develop novel antagonists of oncogenic Activator Protein-1
(Alternative Format Thesis)**

Lathbridge, Alex

Award date:
2020

Awarding institution:
University of Bath

[Link to publication](#)

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

Copyright of this thesis rests with the author. Access is subject to the above licence, if given. If no licence is specified above, original content in this thesis is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC-ND 4.0) Licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>). Any third-party copyright material present remains the property of its respective owner(s) and is licensed under its existing terms.

Take down policy

If you consider content within Bath's Research Portal to be in breach of UK law, please contact: openaccess@bath.ac.uk with the details. Your claim will be investigated and, where appropriate, the item will be removed from public view as soon as possible.

Helical constraints and *in silico* library screening to develop novel antagonists of oncogenic Activator Protein-1

Alexander Lathbridge

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Biology and Biochemistry

August 2019

COPYRIGHT

Attention is drawn to the fact that copyright of this thesis rests with the author and copyright of any previously published materials included may rest with third parties. A copy of this thesis has been supplied on condition that anyone who consults it understands that they must not copy it or use material from it except as permitted by law or with the consent of the author or other copyright owners, as applicable.

This thesis may be made available for consultation within the University Library and may be photocopied or lent to other libraries for the purposes of consultation with effect from.....

Signed on behalf of the Faculty of Science

Contents

Contents	2
Acknowledgements	4
Dedication	4
Abstract	5
List of Abbreviations	6
Chapter 1 - Introduction.....	8
1.1 Structure of the Leucine Zipper	9
1.2 Physiological Role of Activator Protein-1	12
1.2.1 Pathological Role of AP-1.....	15
1.3 AP-1 as an Anti-Cancer Target.....	16
1.3.1 Small Molecule AP-1 Antagonists	17
1.3.2 Peptide Antagonists	18
1.4 Methods for Designing Stable and Specific AP-1 Antagonists	19
1.4.1 <i>Ex silico</i>	19
1.4.2 <i>In silico</i>	21
1.5 Thesis Aims	22
Chapter 2 - Materials And Methods	24
2.1 Peptide Synthesis.....	24
2.1.1: <i>b-f</i> Lactamisation	25
2.2 Circular Dichroism (CD).....	25
2.2.1 Scans	26
2.2.2: Thermal Scans.....	26
2.2.3: Thermal Melts	27
2.3: Size Exclusion Chromatography (SEC).....	27
Chapter 3 - Software Development.....	28
3.1 Library Screen	29
3.1.1 Core Contributions (CoreCalc)	32
3.1.2 Electrostatic Contributions (ESCalc).....	33
3.1.3 Helical Propensity Contributions (HPCalc)	34
3.2 <i>In Silico</i> PCA (isPCA)	34
3.3 <i>In Silico</i> CANDI (isCAN)	37
3.4 Heptad Cassette Predictor (HCAP).....	40
3.5 Library Comparison Tool	42
3.6 Quad Coil Interaction Prediction Algorithm (QCIPA).....	43
3.6.1 Anti-Parallel Coiled Coil Screening	43
3.6.2 Progressive Coiled-Coil Set Expansion	44
Chapter 4 - Computational Competitive and Negative Design to Derive a Specific cJun Antagonist.....	45
4.1 Introduction	47
4.2 Materials and Methods	48
4.3 Results and Discussion	51
4.4 References	56
4.5 Supporting Information	58
4.6 Concluding Remarks on Chapter 4	60
Chapter 5 - Coupling Computational and Intracellular Screening and Selection Towards Co-compatible cJun and cFos Antagonists.....	61
5.1 Introduction.....	63
5.2 Materials and Methods	64
5.3 Results and Discussion	65
5.4 Conclusions	71
5.5 Supporting Information	74
5.6 Concluding Remarks on Chapter 5	78

Chapter 6 - Combining Constrained Heptapeptide Cassettes with Computational Design To Create Coiled-Coil Targeting Helical Peptides	79
6.1 Introduction.....	81
6.2 Results and Discussion	82
6.3 Materials and Methods	90
6.4 References	91
6.5 Supporting Information	93
Chapter 7 - Conclusion	105
7.1: <i>In Silico</i> Screening of Coiled Coils.....	105
7.2: Truncation and Elongation	106
7.3: Summary Conclusions	108
Thesis References.....	109
Chapter A1 - Computational Prediction and Design for Creating Iteratively Larger Heterospecific Coiled Coil Sets	133
A1.1 Introduction	135
A1.2 Materials and Methods.....	136
A1.3 Results and Discussion.....	139
A1.4 Conclusion	144
A1.5 References.....	145
A1.6 Supporting Information.....	147

Acknowledgements

To begin with, I would like to thank Dr Jody Mason welcoming me into his lab. His tenacity and determination to approach big questions from all angles has made for years of rewarding research. The work described within this thesis is the result of years of guidance, encouragement, and advice for which I am extremely thankful.

It goes without saying that the last four years of work would have been made a lot more difficult without my colleagues within the Biology & Biochemistry department and others across at the University of Bath. I am incredibly appreciative of all the insight and aid, both in and out of the lab. Additionally, I am indebted to my friends and family outside of academia and research for reminding me to see the bigger picture.

Finally, none of this would have been possible without the constant support of my partner. This thesis is a testament to their belief in my abilities (unwavering, even in the face of my own uncertainty and self-doubt).

Dedication

To my parents

(Please stop asking if I plan to go to medical school)

Abstract

Activator Protein 1 (AP-1) is a basic leucine zipper (bZIP) transcription factor that is formed from the dimerisation of cJun and cFos proteins to form a coiled coil (CC). As AP-1 is implicated in the pathways of multiple cancers (such as breast cancer, cervical cancer, and colon cancer) this α -helical peptide-peptide interface represents an interesting target for the inhibition of AP-1 formation and transcriptional activity. Peptidomimetic sequences based on cFos are able to preferentially bind to cJun. Three methods were explored to develop and optimise the use of *in silico* screening methods in the design of cJun antagonists. Firstly, *in cellulo* screening techniques were modelled computationally and used to design a stable and specific cJun antagonist. Secondly, this method was further expanded *in cellulo* with Protein-fragment Complementation Assay screening, with the aim of validating the peptide library generated. Both peptides derived from these techniques showed similarly high levels of stability with cJun and promoted instability in negative and off-target complexes. The *in-cellulo* derived sequence showed additional co-compatibility with a previously characterised cFos antagonist and allowed for specific targeting of both AP-1 components. Finally, a set of heptapeptide “cassettes” were synthesised and characterised in both a linear and cyclic form, using *b-f* lactam bridges. The variances in helicity seen in the lactamised forms highlighted differences in constraint tolerance. This data was coupled with computational screening data to modularly select a combination of cassettes predicted to bind to cJun. Full-length peptides selected computationally were shown to have decreased homodimeric and heterodimeric stability, whereas the combined *in-silico/in-vitro* selections showed high levels of stability and were able to successfully target cJun. This provides novel insights into the tolerance of constraints and the design of peptides for stability and helicity. Overall, the methods developed (and the peptides derived) represent a step forward in the design of specific peptide therapies capable of targeting AP-1.

List of Abbreviations

AP-1 – Activator Protein 1

bCIPA - bZIP Coiled-Coil Interaction Prediction Algorithm

BOW – Base Optimised Weights

bZIP – Basic Leucine Zipper

CANDI - Competitive and Negative Design Initiative

CC – Coiled Coil

CD – Circular Dichroism

CE – Cluster Expansion

cFos – cellular-Fos

cJun – cellular-Jun

CLASSY - cluster expansion and linear programming-based analysis of specificity and stability

COPPA – Coiled Coil Peptide Prediction Algorithm

ddH₂O - double-distilled water

DHFR - Dihydrofolate Reductase

DMSO - Dimethyl Sulfoxide

fH – fractional helicity

Fra-1/2 – Fos Related Antigen 1 or 2

FRET - Fluorescence Resonance Energy Transfer

HCAP – Heptad Cassette Predictor

HBS – Hydrogen Bond Surrogates

HPLC - High-Performance Liquid Chromatography

ILP - Integer Linear Programming

isCAN – *in silico* CANDI

isPCA – *in silico* PCA

JDP2 – Jun dimerisation protein 2

JNK - cJun N-terminal kinases

KIHP – knobs into holes packing

K_d - dissociation constant

KPP – potassium phosphate buffer

LZ – Leucine Zipper

MAPK - mitogen-activated protein kinase

MeCN – Acetonitrile

MMP – Matrix Metalloproteinase

MRE – Mean Residue Ellipticity

MTT - 3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyltetrazolium bromide

NFAT - Nuclear factor of activated T-cells

NMR - Nuclear Magnetic Resonance

PCA Protein-fragment Complementation Assay

SEC - Size Exclusion Chromatography

SPPS - Solid Phase Peptide Synthesis

STAT - Signal transducer and activator of transcription

TA - Transactivation

T_m – melting temperature

TF – Transcription Factor

TPA – 12-O-tetradecanoylphorbol-13-acetate

TRE – TPA Response Element

Y2H – Yeast-2-hybrid

CHAPTER 1 - INTRODUCTION

The transcription factor (TF) Activator Protein 1 (AP-1) is a basic leucine zipper (bZIP) formed as a dimeric complex of proteins across the Jun, Fos, ATF and MAF families (Chinenov and Kerppola 2001; Eferl and Wagner 2003). Within the bZIP family of proteins, there exists a shared core architecture of a bZIP domain which is flanked N- or C-terminally by a transactivation (TA) domain (Seldeen et al. 2010).

The bZIP domain itself is composed of 2 separate regions which serve different roles in the transcriptional function of the protein. As shown in Figure 1.1, there is a N-terminal basic sequence which mediates AP-1 recognition and binding to its consensus sequence (GTGACTCA) on DNA – known as the tumour promoting agent (12-O-tetradecanoylphorbol 13-acetate or TPA) response element (TRE) (Angel and Karin 1991). At the C-terminus, the α -helical leucine zipper (LZ) region mediates the dimerisation of two bZIP proteins via the formation of a coiled coil structure.

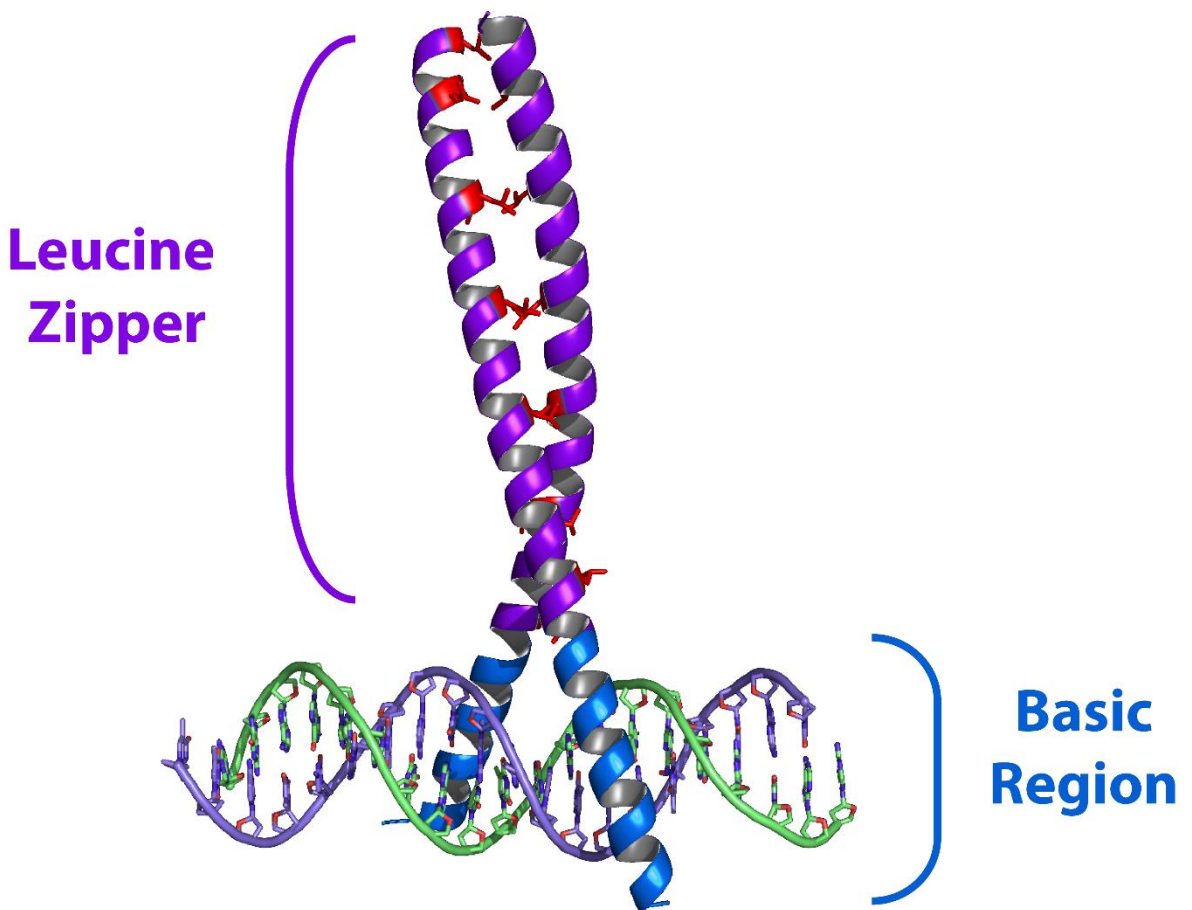


Figure 1.1 Structure of bZIP AP-1 bound to DNA (PDB: 1FOS). The binding of bZIPs to specific DNA sites is facilitated by the (often Arginine-rich) basic region at the N-terminus (blue). The Leucine Zipper region at the C-terminus (purple) mediates the dimerisation of the proteins via the formation of a strip of Leucine residues (red) at every 7th position

In a wider genomic context, the comparison of diversification has shown that the evolution of bZIPs originated from a single eukaryotic protein and the resulting evolution of bZIP family proteins incrementally increased in complexity, while retaining important roles within the regulation of cellular processes (Deppmann et al. 2006; Jindrich and Degnan 2016; Kani et al. 2017). Within humans, there exist 21 bZIP families and over 50 distinct members, with the potential for over 2000 distinct dimers to form and recognise binding sites (Seldeen et al. 2010; Reinke et al. 2013).

The TRE is a consensus sequence found within transcriptional regulatory regions on DNA. It was first discovered in the promoters of metallothionein IIa (hMTIIa) and simian virus 40 (SV40). These sites facilitate binding to specific structures on proteins. Within AP-1, the structure of the basic region allows the residues to make contacts that are base-specific within the major groove of DNA and the flexibility of the LZ region means that AP-1 does not exclusively bind to a specific orientation of the TRE (Glover and Harrison 1995).

Depending on the constituent bZIP proteins, AP-1 can form in many different complexes. The Jun family of proteins (JunB, JunD, and cJun) are able to form transcriptionally active AP-1 via both homodimerisation and heterodimerisation with Fos family proteins (Darlyuk-Saadon et al. 2012). Conversely, Fos family proteins (Fra1, Fra2, FosB, cFos) display far less flexibility and require heterodimerisation with Jun family proteins to form AP-1, with this heterodimer able to bind DNA tighter than the Jun homodimer (Ransone et al. 1993). Given the diversity of roles that they play physiologically, the structurally mediated selectivity and specificity of bZIP formation is of key importance across all families – including that of Jun and Fos (Ransone et al. 1993). Relative to simpler eukaryotes, the greater ability to form heterodimers and engage with distinct binding sites suggests that the multitude of possible complexes contributed to the further organismal complexity (Lamb and McKnight 1991; Reinke et al. 2013; Rodríguez-Martínez et al. 2017).

1.1 Structure of the Leucine Zipper

The leucine zipper is a short α -helical region of typically 30-40 residues that mediates binding between bZIP proteins via α -helical interactions (Crick 1953; O'Shea et al. 1991). The general structure of the LZ incorporates architecture based on a sequence of residues conforming to a seven residue pattern known as a heptad repeat. The heptad pattern, repeating for the length of the LZ is defined as **(*abcdefg*)_n**, with each position populated by characteristically different residues.

This sequence forms a coiled coil owing to its α -helical properties. This CC is a common structural motif and it exists in 3 – 5% of amino acids within proteins and typically consist of 2 – 5 α -helices wrapped around one another in a parallel or anti-parallel manner (Mason and Arndt 2004). Canonically, an α -helix exhibits 3.6 residues per turn of the helix but within

the CC, this decreases this to 3.5 – allowing for a complete repeat of the heptad every 2 turns. When discovered by Crick, this was geometrically described as a 20° incline between the axes of the helices, which then allows the CC itself to follow a gradual helix with a pitch angle of ~10° (Crick 1953). This gives rise to the formation of the supercoiling of the structure, seen as a shallow left-handed helical structure, with the presence of a “strip” of hydrophobic residues embedded in the core of the heptad for the entire length of the CC. Most of the interhelical interactions within the LZ structure – influencing stability and specificity- involve residues in positions **a**, **d**, **e**, and **g** (Alber 1992).

The structural stability of the CC is achieved via “knobs into holes packing” (KIHP), in which the side-chains of residues of position **a** and **d** (“knobs”) pack into the spaces on the opposing helix (“holes”). Within position **a**, the knob is packed into a hole surrounded by the side chains of residues at position **a**, position **g**, and those of two position **d**. This is mirrored for the knob on position **d**, with its hole surrounded by side chains of positions **d**, **e**, and two of position **a** (Crick 1953; O’Shea et al. 1993; Glover and Harrison 1995). The “ridges into grooves” packing (RIGP) model is similar, with interfacial interactions between the two helices incorporating the reciprocal placement of side chains (Chothia et al. 1977; Chothia et al. 1981). Unlike KIHP, RIGP defines the environment for single residue side chain placement as a “groove” on the opposite helix defined by 2 residues, rather than the 4 involved in the defining of the “hole” in KIHP (Efimov 1999).

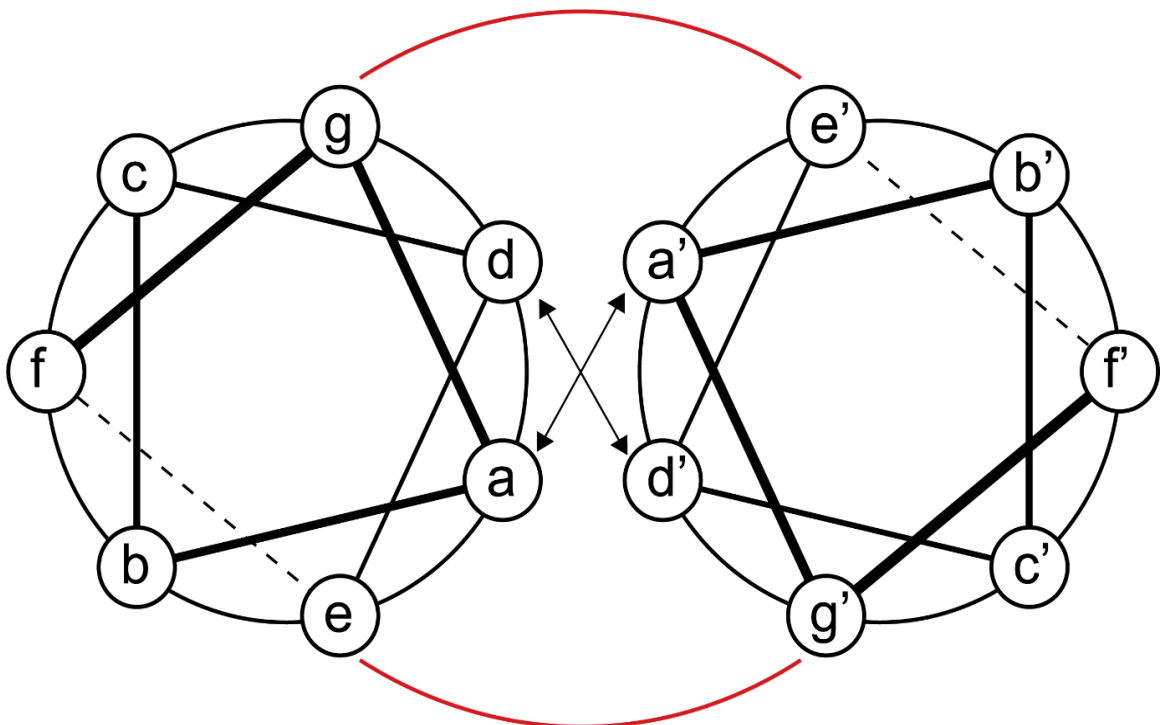


Figure 1.2 Helical wheel showing the relative positions of residues within the coiled coil. Residues at positions **a** and **d** form the buried hydrophobic interface, with potential electrostatic interactions (red) formed between residues at positions **e** and **g**.

As shown in Figure 1.2, positions **a** and **d** form the buried core region of the zipper, most frequently populated by hydrophobic residues, such as leucine, isoleucine, valine and alanine. Position **d** is typically observed to be a leucine residue and one reason suggested for this phenomenon is the primary role that the leucine repeat might play in the identification of bZIP proteins (O'Shea et al. 1991). This leucine repeat at every 7th residue is the factor which contributes the most to the stabilisation of dimerisation between the LZs from **d-d'** interactions, with the **a-a'** interactions. The conservation of leucine at **d** is thought to suggest that the interfacial packing and CC stability would be lost if the residue were replaced with one smaller, larger, or unbranched (Hu et al. 1990; Hodges et al. 1990; Hu and Sauer 1992).

Due to the precision of the KIHP geometry, it's known that the residues within the sequence affect the oligomeric state; with the understanding that different combinations of side-chains influence how the helices wrap around one another (Woolfson and Alber 1995; Nautiyal et al. 1995). This sequence-to-structure understanding has shown that residues at these core positions are linked to the oligomeric states of the bZIPs, with isoleucine at **a** and leucine at **d** promoting parallel dimers. (Harbury et al. 1993) The selection of isoleucine at both **a** and **d** has been shown to promote a trimeric configuration, and tetramers promoted with an **a/d** selection of leucine and isoleucine respectively.

Within bZIPs generally, the flexibility of **a** positions to accommodate polar residues allows the specificity, oligomeric state, and helical orientation to be controlled (Potapov et al. 2015). This location benefits from the ability to form intrahelical hydrophobic interactions with leucine residues at **d** and this is energetically favours β -branched amino acids valine and leucine (Acharya et al. 2006). The selection of charged or polar residues within the core is not uncommon within AP-1, with threonine and lysine present at the **a** position of cFos and asparagine at the same position on cJun. Structurally, the side chain of the basic residue lysine in cFos packs against its main chain as well as against those of side chains within positions **a** and **g** in cJun, conforming to the KIHP pattern (Glover and Harrison 1995).

Similarly, the selection of asparagine in cJun is not uncommon in CCs. Although its presence within the core is destabilising, the trade-off is that it is thought that asparagine-asparagine inter-helical interactions within CCs limits the oligomeric state to that of a dimer, with asparagine -> valine mutants exhibiting mixtures of dimeric and trimeric structures (O'Shea et al. 1991; Hartmann et al. 2009; Fletcher et al. 2017; Thomas et al. 2017). Within the cJun-cFos interaction, the asparagine within cJun is able to donate a hydrogen bond to the preceding glutamate within cFos due to the solvent exposed nature of asparagine's side chain (Glover and Harrison 1995). Additionally, the presence of polar residues at **a** has been shown to specify the orientation and alignment of the CC. The interaction of neighbouring polar residues ensures that the CCs align correctly to ensure correct interaction between residues at positions **g** and **e** (Alber 1992).

In parallel dimeric CCs, the residues at the **e** and **g** positions form attractive or repulsive electrostatic $\mathbf{g}^i\text{-}\mathbf{e}^{i+1}$ interactions between the helices (O'Shea et al. 1991). Whereas the core interactions are thought to contribute primarily to stability, these interactions contribute to specificity in multiple forms – helix orientation, oligomeric state, and the formation of either heterodimers or homodimers (Graddis et al. 1993; Monera et al. 1993; Zeng and Zimmerman 1997). As aforementioned, it is thought that these **g** and **e** residues also play a role in the KIHP structure, with their generally long and charged side chains forming the solvent-exposed wall of the hole and the aliphatic regions packing against the knob within (Glover and Harrison 1995; Havranek and Harbury 2003).

Residues at positions **b**, **c**, and **f** constitute the almost completely solvent-exposed backbone “outerface” of the CC. With limited data on the matter, the role that residues at these positions play is poorly understood (Baxter, Ullman, et al. 2017). Due to their position away from the binding interface, it is thought that these residues promote stability and solubility of the α -helix (O'Neil and DeGrado 1990; Kaplan et al. 2014). This contribution to helical propensity is reflected in *in silico* prediction models (Mason et al. 2006). Early characterisation of the CC highlighted the ability of **e** and **g** residues to form ionic interactions with residues at **b** and **c**, with additional studies suggesting potential intrahelical interactions between **b-f** and **f-c** (O'Shea et al. 1991; Mason et al. 2006).

1.2 Physiological Role of Activator Protein-1

The main role of AP-1 TFs is to convert extracellular signals into intracellular expression changes of specific genes. The activity of the transcription factor is modulated by interactions with other transcriptional regulators and is linked to various signal pathways by kinases positioned upstream – with all mitogen-activated protein kinase cascades playing a role in this process (Rincon and Flavell 1994; Jochum et al. 2001; Shaulian 2010).

AP-1 TFs regulate cellular proliferation by regulating the expression of genes vital to the cell cycle (Garces de Los Fayos Alonso et al. 2018). This is observed in cJun's regulation of the cell cycle to promote progression by stimulating cyclin D1 (as shown in Figure 1.3) while also inhibiting the p53 tumour suppressor (Oien et al. 2000). The latter results in the reduction of cyclin-dependent kinase inhibitor p21 and the progression of the cell cycle from G1 to S phase is promoted (Schreiber et al. 1999). This has a reciprocal connection with JunB, which is able to slow the progression of the G1 to S phase transition through repression of the cyclin D1 promoter and the transcriptional activation of p16^{INK4a} (Latifa Bakiri, Dominique Lallemand et al. 2000; Passequé and Wagner 2000). cJun also exhibits tumour suppressive properties identified in studies of cJun deficient cells. These cells were found to suffer spontaneous DNA damage (and therefore early senescence) – implying that cJun potentially has a role in the DNA repair pathway (MacLaren et al. 2004).

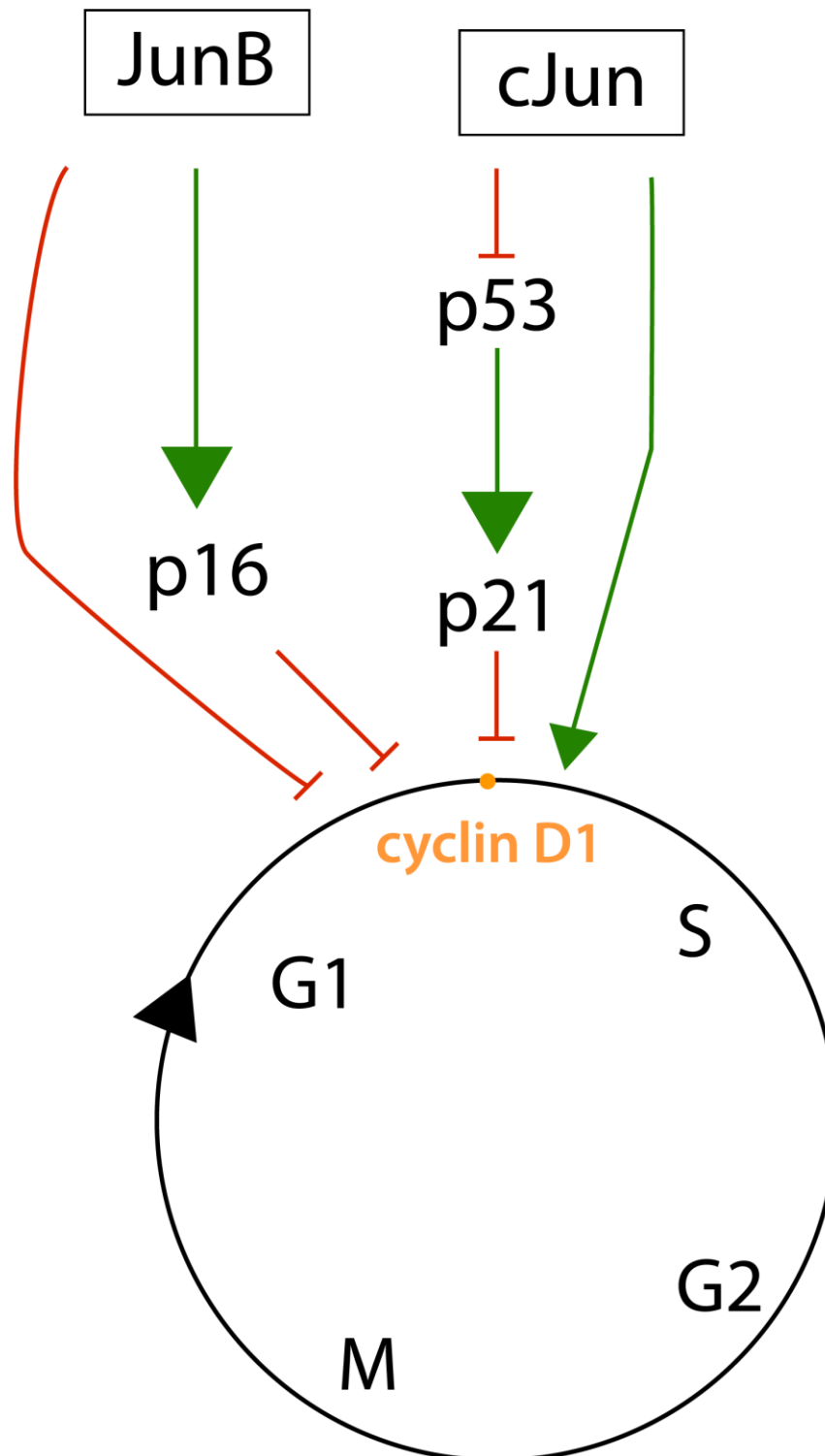


Figure 1.3 cJun and JunB have roles within the cell cycle, notably at the G1-S transition via Cyclin D1 regulation. By inhibiting the tumour suppressor p53 and direct stimulating Cyclin D1, cJun promotes the transition. Conversely, direct repression of the Cyclin D1 and activation of p16^{INK4a} slows this transition.

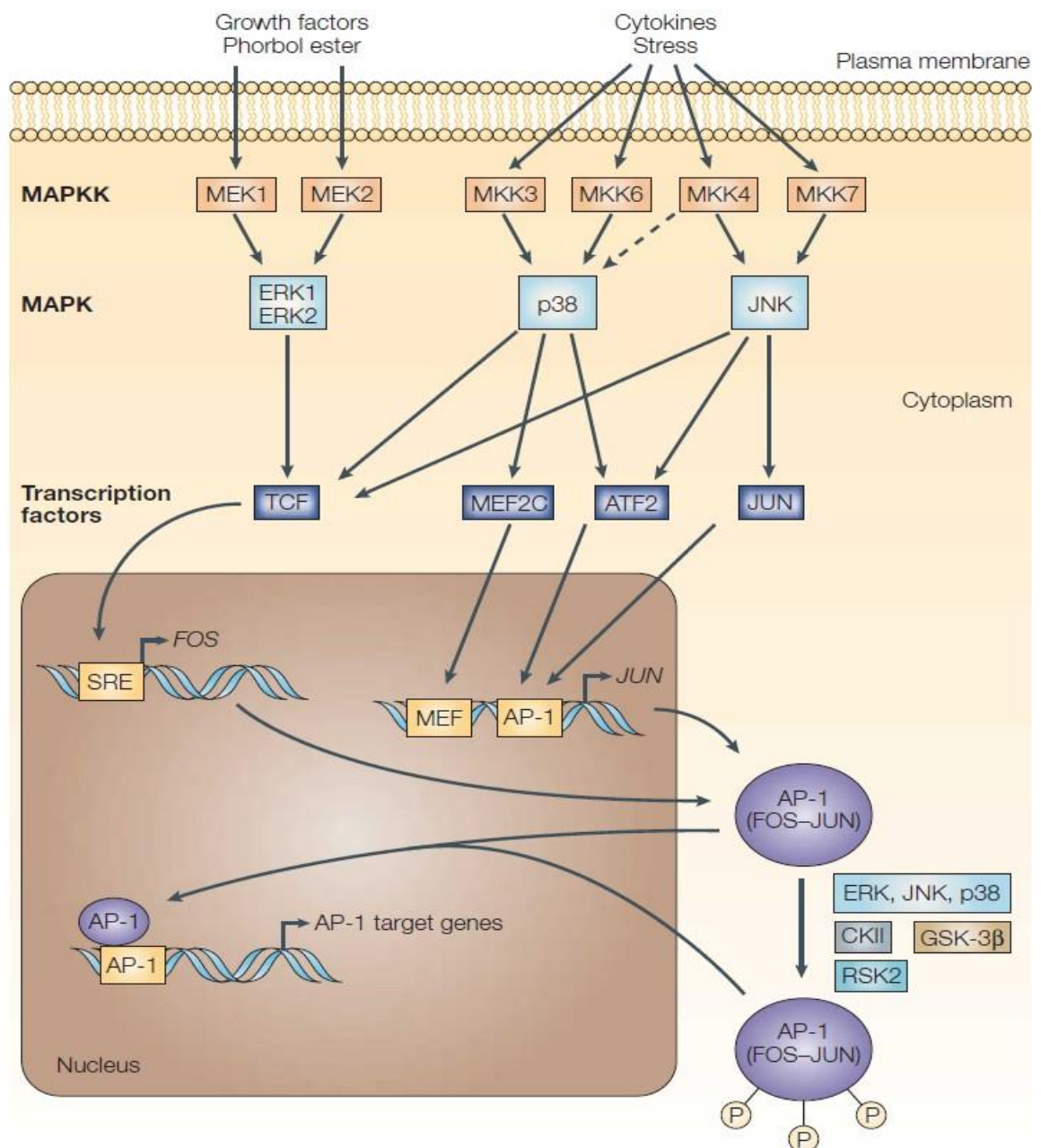


Figure 1.4 AP-1 expression and activity is primarily regulated by multiple MAPK pathways, including the JNK pathway. These signalling pathways are activated by a number of external signalling mediators, including cytokines, stresses, hormones, and growth factors. The increase in constituent cFos and cJun promotes formation of AP-1. The hyperactivity of these signalling pathways can lead to the dysregulation of AP-1 and promote tumourigenesis through the dysregulation of its target genes (including the upregulation of constituent cJun). Figure adapted - (Eferl and Wagner 2003)

As shown in Figure 1.4, expression of cJun and its transcriptional activity is regulated primarily by MAPK pathways, with a prominent one being the cJun amino terminal kinase (JNK) pathway – which phosphorylates and modulates Jun proteins (Dérjard et al. 1994; Kyriakis et al. 1994). This pathway is activated by a number of external mediators such as tumour necrosis factor α (TNF- α), a cytokine which is critical in immune response, proliferation, differentiation, and apoptosis (Chen and Goeddel 2002; Ventura et al. 2003).

The cellular expression of cFos is transient, with expression occurring as part of the nuclear response to stimuli including neurotransmitter release, growth factors, and sensory stimulation (Curran and Morgan 1987; Caputto et al. 2014). With the cessation of the stimulus, the half-life of cFos is within the hour range, with degradation of the protein occurring almost exclusively via the proteasome (Basbous et al. 2008). Due to this rapid nuclear turnover, a method of translational targeting first described in the TF c-myc has been discussed – with a mechanism directing cFos mRNA to the perinuclear cytoplasm to facilitate efficient transport to the nucleus (Caputo et al. 1986; Hesketh et al. 1994; Dalgleish et al. 2001).

The physiological roles of AP-1 proteins have been probed via mouse model gene disruption studies, which have highlighted a wide variety of cellular pathways and tissues affected by their presence or lack thereof. In many cases, AP-1 proteins contribute to the viability of mice - the loss of JunD is linked to male sterility, and embryonic lethality is observed with the knockout of Fra1, JunB, or cJun (at 9.5, 10, and 10.5 weeks respectively). It is thought that the MAPK and AP-1 proteins are linked to the degradation of the extracellular matrix (ECM) in foetal membranes (Lappas et al. 2011). Their increased expression at supracervical sites is thought to play a role in increasing the susceptibility of foetal membrane rupture susceptibility through matrix metalloproteinase-9 (MMP-9) modulation (Sato and Seiki 1993; Chakraborti et al. 2003; Yonemoto et al. 2006; Sitras et al. 2008). Analysis of mice lacking cFos has shown that, while fertile, viable, and able to survive until adulthood, they are osteopetrotic due to a lack of osteoclasts (Wang et al. 1995; Passequé and Wagner 2000; Hess 2004). The loss of FosB is linked to a range of neurobehavioral dysfunctions in postpartum mice, including changes in emotional behaviour and a decrease in both nurturing and infanticide behaviour (Kuroda et al. 2008). There is an element of overlap in the functions of AP-1 proteins, with various proteins able to partially cover the proteins that have been knocked out. Whereas Fra1 is able to rescue from the osteopetrotic phenotype associated with a loss of cFos, the latter is not able to rescue from the lethality associated with the loss of Fra1 (Schreiber et al. 2000). Despite the opposing roles within the cell cycle, a knock-in of JunB (and JunD) rescues embryonic lethality of a cJun knockout until birth (Passequé and Wagner 2000).

1.2.1 Pathological Role of AP-1

Due to the intricate and complex role of AP-1 within cellular and physiological events, deregulation of mechanisms involving the TF can have a multitude of pathological outcomes, including several cancers, organ damage, fibrosis, asthma, psoriasis, and other inflammatory conditions (Xanthoudakis et al. 1994; Chinenov and Kerppola 2001; Zenz et al. 2007; Trop-Steinberg and Azar 2017; Gungl et al. 2018; Sun et al. 2019).

In several studies, AP-1 proteins have been shown to play an important role in cardiovascular disease. JunD has been found to have a role in the regulation of heart growth and protection from cardiac apoptosis (Ricci et al. 2005). It is the only Jun family protein expressed highly within the mammalian heart and regulates AP-1 activity to protect against stress-induced hypertrophic growth in cardiomyocytes (Hilfiker-Kleiner et al. 2005). Both JunB and FosB are involved in the response to heart injury by the regulation of MMP-2, which influences ventricular performance (Alfonso-Jaume et al. 2006). Fra2 expression (predominantly in the lung) is one of the contributing factors in the development of pulmonary fibrosis by its regulation of fibrogenic mediators (Hardie et al. 2009; Ucer0 et al. 2019). Fra1 is involved in the modulation of proinflammatory and profibrotic genes in order to protect tissue from injury, with Fra1 deficient mice displaying greater levels of fibrosis (Rajasekaran et al. 2012).

Inflammatory diseases involve the dysregulation of the immune system and can result in overexpression of cytokines, enzymes, and immune cells involved in the inflammatory response. AP-1 (along with NFAT and STAT proteins) is considered to be a pro-inflammatory TF that regulates the expression of tumour necrosis factor α and other cytokines (Uluçkan and Wagner 2016). These cytokines modulate the recruitment and activation of immune cells, with chronic inflammation linked to disorders such as asthma, inflammatory bowel disease, and psoriasis (Zenz et al. 2007; Trop-Steinberg and Azar 2017)

1.3 AP-1 as an Anti-Cancer Target

Estimated to account for 20% of oncogenes, the deregulation of TF activity within signalling pathways characterises the majority of human cancers (Vaquerizas et al. 2009; Yeh et al. 2013). Changes in cellular differentiation, apoptosis, and other processes are co-ordinated and regulated via TFs (Mukherjee et al. 2018). Both cJun and cFos were initially discovered as cellular counterparts to retroviral oncoproteins v-Jun and v-Fos found in avian sarcoma virus 17 (ASV17) and Finkel-Biskis-Jinks murine sarcoma virus (Maki et al. 1987; Watanabe et al. 2002). Other AP-1 family member proteins have also been shown to be oncoproteins (JunB, JunD, Fra1, Fra2, and FosB)

The presence of AP-1 (and AP-1 proteins) in oncogenic pathways can be attributed to two main factors. Firstly, due to the previously described semi-redundant nature of the family, AP-1 proteins have overlapping and non-overlapping functions which highlights the importance of specificity and composition in the function of the TF. AP-1 has been described as a “double-edged sword” due its implication in cancer-related pathways as well as the observed ability of its proteins to display both tumour suppressant and pro-oncogenic behaviour depending on the physiological context (Eferl and Wagner 2003). The second

factor is that AP-1 sits at key intersecting points of signalling pathways for the transcriptional regulation of many genes – including potential oncogenes (p53 – as aforementioned – plays a crucial role in the cell cycle).

AP-1 is implicated in the pathogenesis of cancers including breast cancer, skin cancer, colon cancer, hepatocellular carcinoma, cervical cancer, human oral squamous cell carcinoma (Milde-Langosch 2005; Apostolou et al. 2013; Eckert et al. 2013; Kappelmann et al. 2014; Chen et al. 2017; Tyagi et al. 2017). Within the dysregulated oncogenic signalling networks, AP-1 proteins have been found to contribute to the control of apoptosis, survivability and resistance to interventions (Zenz et al. 2007; Tyagi et al. 2017; Endo et al. 2018). This is highlighted in the genes that they modulate. cJun upregulates genes linked with invasiveness, angiogenesis, and proliferation while downregulating genes stimulating apoptosis and (as with JunD) inhibiting proliferation. (Lamb et al. 1997; Park et al. 1999; Schreiber et al. 1999; Rebollo et al. 2000; Whitfield et al. 2001; Ivanov et al. 2001; Miao et al. 2017). Similarly, cFos upregulates genes involved in the methylation of DNA, the stimulation of apoptosis, angiogenesis and (as with Fra1) invasiveness (Hennigan et al. 1994; Jooss and Müller 1995; Kasibhatla et al. 1998; Bakin and Curran 1999; Marconcini et al. 1999). cJun (and JunB) also have potential tumour suppressing behaviour and can downregulate genes that inhibit apoptosis and proliferation.

Within prostate cancer, the upregulation and increased activity of AP-1 proteins is associated with tumourigenesis, metastasis, angiogenesis and invasion (Leach et al. 2017; Udayappan and Casey 2017). Additionally, these proteins have been shown to confer resistance to castration-based interventions and its role in proliferation has been targeted through the use of finasteride (Babu et al. 2013; Wang et al. 2017). In epithelioid hemangioma (an aggressive vascular neoplasm found in soft tissue and bones) Fos mutations have been identified as potential drivers of local angiogenesis through the Notch signal pathway and the production of MMPs (Eferl and Wagner 2003). Truncation at the C-terminus has been shown to encode a highly stable isoform of cFos (with a half-life in excess of 8 h) that is resistant to ubiquitin-independent degradation by the 20S proteasome due to the loss of a helical motif, allowing prolonged endothelial sprouting (Van IJzendoorn et al. 2017).

1.3.1 Small Molecule AP-1 Antagonists

Traditional therapies to target cancer make use of small molecules. These are low weight (≤ 500 Da) and have a small surface area for interaction. The druggability of many targets is largely based on the ability of the small molecule to form a complex by maximising contact on the surface area. In many cases, this is made possible by the presence of a hydrophobic pocket on the target macromolecule. This form of chemotherapy lacks specificity towards

cancer cells and can result in the regrowth of resistant tumour clones (Ellert-Miklaszewska et al. 2017).

AP-1 proteins feature interaction surfaces with structure unsuited for conventional targeting (Darnell 2002). As such, small molecule inhibitors have shown limited success and the development of specific therapies has proven to be challenging (Che et al. 2006; Ye et al. 2014; Papavassiliou and Papavassiliou 2016)

The potential use of natural bioactive agents as AP-1 antagonists has been extensively studied (van Dam and Castellazzi 2001; Kappelmann et al. 2014; Ibrahim et al. 2018). The diversity and specificity of the AP-1 complexes has made identifying molecules able to target AP-1 interactions difficult, with many of these indirectly modulating the pathways controlling AP-1 activity (Tewari et al. 2018). Resveratrol, a flavonoid polyphenol, has been reported to negatively regulate AP-1 via the inhibition of the JNK-cJun/AP-1 signalling cascade. However, many of the natural agents studied modulate AP-1 activity via a multitude of highly disruptive mechanisms, including the suppression of MAPK signalling cascades and the prevention of DNA binding (Dedieu and Lefebvre 2006; Malorni et al. 2016).

Only one inhibitor for AP-1 has entered human clinical trials, with many of the compounds available lacking specificity and targeting other TFs. A small cyclic compound, T-5224 was investigated in Phase II human clinical trials as an inhibitor which displayed a degree of selectivity over other TFs but was not further studied (Uchihashi et al. 2011; Garces de Los Fayos Alonso et al. 2018).

1.3.2 Peptide Antagonists

In recent years, greater focus has been given to the development of synthetic peptides, with the aim of characterising the rules governing protein-protein interactions (Crooks et al. 2016). This advancement towards the development of *de novo* sequences has a wide range of potential applications in self-assembling biomaterials, drug delivery, the formation of complex architecture, and other building blocks for assembling synthetic systems (Shlizerman et al. 2010; Gradišar et al. 2013; Yeates et al. 2016; Beesley and Woolfson 2019; Majerle et al. 2019).

Concurrently, peptides have begun to rise to the challenge of targeting proteins that were previously thought to be undruggable (Thansandote et al. 2015; Rastogi et al. 2019). Short peptides have shown high binding affinity to their targets as they are capable to interacting on the flat and extensive interfaces that mediate many protein-protein interactions. Historically, peptide-based therapeutics were discounted during drug design due to the

inherent limitations such as lack of specificity, susceptibility to enzymatic degradation, poor delivery, and inadequate transport through membranes (Vagner et al. 2008).

It is thought that synthetic peptidomimetics could potentially represent a mid-point between small molecules and antibodies, with the bioavailability and permeability of the former and the specificity of the latter (Bruzzoni-Giovanelli et al. 2018; Farley et al. 2019). The flexibility of peptidomimetics in drug design is due to the diversity of natural (and non-natural) amino acids and chemical modifications possible to create more “drug-like” molecules (Tapeinou et al. 2015). These include alterations to the side-chain and backbone, including conjugation, truncation, and helix constraints (Crooks et al. 2011a; Baxter, Perry, et al. 2017).

The use of these constraints in the design of peptide mimetics has been explored to address the issues related to specificity, bioavailability, permeability, and degradation (Craik et al. 2013; Hill et al. 2014; Thansandote et al. 2015; Hoang et al. 2015). This involves the cyclisation of two amino acids to promote the adoption of the α -helix by reducing the entropic cost associated (Rao et al. 2013). Common modalities include hydrocarbon “staples”, hydrogen bond surrogates (HBS), and lactam bridges (Taylor 2002; Patgiri et al. 2008; Mason 2010; Aihara et al. 2015; Walensky and Bird 2015; Hoang et al. 2016; Rezaei Araghi et al. 2018).

The use of HBS to replace backbone hydrogen bonds with covalent bonds offers stability with potentially less interference to the residues involved in protein-protein interaction (Chapman et al. 2004). These techniques have been shown to be increasingly successful in disrupting the protein-protein interactions required to permit transcriptional activity (Staber et al. 2007; Cumaraswamy et al. 2014; Ball et al. 2016; Rezaei Araghi et al. 2018).

1.4 Methods for Designing Stable and Specific AP-1 Antagonists

1.4.1 *Ex silico*

The transcriptional function of AP-1 is dependent on the ability to form as a dimer – with the dimerisation controlled by the LZ domain. As such, many strategies for targeting AP-1 are linked to the rational design of peptide antagonists to target this region with engineered stability and specificity.

This has been guided by structural data, allowing for further understanding into the structural basis of the protein-protein interactions. X-ray crystallography and Nuclear Magnetic Resonance (NMR) studies into LZ formation have highlighted key residues within interactions. The structural models of LZs such as GCN4 (O’Shea et al. 1991) and AP-1 (Glover and Harrison 1995) provided specific insight into the protein-protein interactions of

LZs but it is not currently possible to solely rely on structural data as a form of screening for peptide antagonists.

The Protein Fragment Complementation Assay (PCA) is an example of an *in cellulo* method for the identification of stable protein-protein interactions based on the ability to selectively reconstitute a functional third protein (Pelletier et al. 1998). The bait and the prey proteins are both covalently fused to a fragment of a reporter protein. The low affinity of the two fragments for one another ensures that only the interaction of the bait and prey proteins leads to the re-formation of the reporter protein (Remy et al. 2002). This requires the reporter protein to be able to reconstitute itself without covalent bond formation between the fragments and function to produce a quantifiable output. Dihydrofolate Reductase (DHFR) is an enzyme necessary for cellular survival/growth in media lacking in complex nutrients. Its role as a reporter protein is possible due to the use of Trimethoprim (an inhibitor of DHFR) coupled with the transformation of bacteria with split murine DHFR fused to the prey and predator proteins (Pelletier et al. 1998; Michnick et al. 2000; Remy and Michnick 2015).

This method of protein-protein interaction screening excludes false positives as only the bacteria that express interacting proteins (and form functional DHFR) can grow in the media. The advantage of this is that it does not require the activation of a downstream response. This is the case with a similar technique, Yeast 2 Hybrid (Y2H) screening (Young 1998). Y2H relies on the protein-protein interaction effecting transcription of a downstream reporter gene and is therefore dependent on the restored function of the transcription factor (Joung et al. 2000).

PCA can be extended through the incorporation of negative design (Mason, Müller, et al. 2007). The competitive and negative design initiative (CANDI) introduces competitor sequences to the assay. Cells are unable to grow when a competitor is in complex with the library member of the target (due to the lack of a DHFR fragment). Due to this, each additional competitor sequence represents additional stringency with the potential formation of 2 additional off-target complexes. This contrasts highlights the difference between CANDI and techniques relying on library-target affinity, the latter of which are able to result in off-target complexes of increased stability (Grigoryan et al. 2009b). However, it has been shown that in some systems only marginal differences exist between the binding affinity of desired complexes and that of the off-target complexes (Mason et al. 2006; Mason, Müller, et al. 2007).

This use of bait and prey proteins is not unique to PCA or Y2H. Fluorescence resonance electron transfer (FRET) studies attach fluorophores to both proteins, where one fluorophore's emission wavelength and the other's excitation wavelength are identical (Pollok and Heim 1999; Chen et al. 2007). The bait and prey proteins being in close

proximity to one another results in lower FRET efficiency and this is used as an indicator of binding affinity (Chen et al. 2007; Szalóki et al. 2015).

PCA can also be combined with CIS display – a technique that uses components of bacterial transcription and translation machinery to fuse peptide library members onto specific regions of DNA (Odegrip et al. 2004; Mathonet et al. 2011). This method can be used to screen expansive libraries (10^{10} members) and can select for stringency with truncated and non-truncated versions of the same sequence serving as targets (Baxter, Ullman, et al. 2017). Incorporating the successful members of CIS into PCA allows for screening at both the *in vitro* and *in cellulo* level.

1.4.2 *In silico*

With the increase in structural, stability, and affinity data, the range of computational design approaches targeting bZIPs has expanded. The characterisation of CCs from structure has been explored with computational approaches, with early applications elucidating the principles governing the oligomeric state of the CC (Woolfson and Alber 1995). This has led to identification and filtration of CC structures from RCSB Protein Data Bank (Berman et al. 2000) by searching for packing and features indicative of KIHP (Walshaw and Woolfson 2001), and the use of molecular dynamics to compute binding affinity (Zuo et al. 2012). Further understanding of the structural principles has recently led to the development of CCBUILDER – a tool which allows for a structural model of novel CCs to be built from sequence data (Wood and Woolfson 2018). This structure-based approach to recognising CCs has informed the prediction of CC interactions from sequence. Unlike larger proteins with complex globular structures, the relative simplicity of the CC has allowed LZ sequence data to inform the design of antagonists.

Initial work combined sequence data and experimental data to binarily predict the formation of a CC interaction (Singh and Kim 2001). This focused on the prediction of helical alignment and if a pairwise heterodimeric CC interaction was preferred (instead of two possible homodimeric interactions). This made use of a support vector machine (SVM) approach – a supervised learning model used to classify objects into groups depending on their characteristics (Burges 1998). These characteristics are converted into variables and which have weight vectors applied and optimised through training. These characteristics were developed from previous biophysical data, with the datasets used to train and validate the algorithm containing sequences known either to form CCs or not form CCs (Parry et al. 1977; Parry 1982; O'Shea et al. 1991; Krylov et al. 1998a). It was found that 90% of incorrect partners were able to be eliminated from ~95% of the CC regions found in the dataset. SVM's ability to recognise patterns has been utilised as the basis of many algorithms used in computational biology, including the characterisation of protein-DNA interactions in the identification of TF binding sites (Shameer et al. 2010; Hu et al. 2019). SVM has also been

utilised to develop the Base Optimised Weights (BOWs) system that uses weight vectors and the interhelical pairings of residues at certain positions to predict the formation of CCs for bZIP pairings (Fong et al. 2004). This model assumes that the partnering of CCs is governed by several interhelical interactions at specific positions within the heptad: $\mathbf{a}^i\mathbf{a}'^j$, $\mathbf{d}^i\mathbf{d}'^j$, $\mathbf{a}^i\mathbf{d}'^j$, $\mathbf{d}^i\mathbf{a}'^j$, $\mathbf{d}^i\mathbf{e}'^j$, $\mathbf{g}^i\mathbf{a}'^{i+1}$, and $\mathbf{g}^i\mathbf{e}'^{i+1}$.

Instead of weight vectors, the bZIP CC Prediction Algorithm (bCIPA) is driven by coupling energy data from double-mutant analysis studies in CCs (Krylov et al. 1998a; Acharya et al. 2006). Trained on thermal stability data of 57 dimers, bCIPA removes several of the interactions scored by the BOW model. Instead, it relies on the contributions of $\mathbf{a}\mathbf{a}'$, $\mathbf{d}\mathbf{d}'$, $\mathbf{g}^i\mathbf{e}'^{i+1}$, and $\mathbf{e}^i\mathbf{g}'^{i+1}$ interactions, as well as the contribution of each residue to overall helicity in order to predict the melting temperature (T_m) of the CC (Williams et al. 1987; Mason et al. 2006; Hagemann et al. 2008). Due to its training on natural CCs, bCIPA's accuracy has been shown to decrease when applied to truncated CCs with fewer interactions contributing to scoring (Bromley et al. 2009). When applied to microarray analysis data from 59² bZIP interactions, bCIPA was able to correctly identify 92% of non-interactors and 92% of strong interactors (with BOW predictions reporting 89% and 83% respectively). The incorporation of negative design screening allowed for the discovery of motifs and intramolecular residue interactions that play a role in interaction specificity (Havranek and Harbury 2003). This has been explored further using the cluster expansion and linear programming-based analysis of specificity and stability (CLASSY) – a machine-learning based framework (Grigoryan et al. 2009b). Cluster expansion (CE) searches for low-energy configurations of sequences to determine the energies of CC formation (Zhou et al. 2005). This is combined with integer linear programming (ILP), which models the system as an optimisation problem to be solved. This has been used to optimise target affinity-based peptide designs into sequences that represent a trade-off between stability and specificity in targeting LZs. (Reinke et al. 2010; Chen et al. 2011; Potapov et al. 2015).

1.5 Thesis Aims

Approaches to design and screen antagonists of AP-1 are numerous and diverse. This thesis aimed to identify and develop novel computational techniques for the design of peptide-based antagonists. Additionally, it aimed to develop novel synthetic peptide “building blocks” frameworks, incorporating *in silico* design and helix constraints.

Chapter 3 explores the development of computational techniques used during this thesis, the rationale behind the design, and how the initial toolset allowed for modular expansion with additional scenarios for application.

Chapter 4 explores the expansion of previous *in silico* screening techniques to mimic *in cellulo* methods. The screening of large libraries based on template LZs results in the

selection of an antagonist capable of maximising desired complex stability with concurrent minimisation of off-target complex stability.

Chapter 5 explores the enhancement of this *in silico* screening through the addition of PCA. This allows for increased stringency of the screening, with a smaller library expressed on PCA than previous designs on the same template. This results in a successful AP-1 antagonist that has co-compatibility with another antagonist and a system which is capable of binding both AP-1 proteins.

Chapter 6 focuses on furthering our understanding of helical constraints in CCs and the use of modular design in the CC engineering. This study compares lactamised and non-lactamised sequences designed *in silico* and *in vitro* to explore the effects of heptad stability in the tolerance of helical constraints and their effect on full-length sequences. The results highlight the ability to develop stable truncated CCs that are able to bind to cJun with varying affinity, dependent on the stability of the heptads incorporated into the full-length sequence. This represents an advancement into understanding the rules governing constant tolerance and the development of heptad libraries able to promote the formation of desired complexes.

CHAPTER 2 - MATERIALS AND METHODS

2.1 Peptide Synthesis

Rink amide ChemMatrix™ resin was obtained from PCAS Biomatrix, Inc. (St.-Jean-sur-Richelieu, Canada); Fmoc L-amino acids and 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate or benzotriazol-1-yl-ox-ytritypyrrolidinophosphonium hexafluorophosphate were obtained from AGTC Bioproducts (Hessle, UK); all other reagents were of peptide synthesis grade and obtained from ThermoFisherScientific (Loughborough,UK).

Peptides were synthesised on a 0.1-mmol scale on a PCAS ChemMatrix™ Rink amide resin using a Liberty Blue™ microwave peptide synthesiser (CEM; Matthews, NC) employing Fmoc solid-phase techniques with repeated steps of coupling, deprotection and washing (4 × 5 ml dimethylformamide).

Coupling was performed as follows: Fmoc amino acid (5 eq), 2-(1H-benzotriazole-1-yl)-1,1,3,3 tetramethyluronium hexafluorophosphate or benzotriazol-1-yl-oxytritypyrrolidinophosphonium hexafluorophosphate(4.5 eq) and diisopropylethylamine (10 eq) in dimethylformamide (5 ml) for 5 min with 35-W microwave irradiation at 90 °C.

Deprotection was performed as follows: 20% piperidine in dimethylformamide for 5 min with 30-W microwave irradiation at 80 °C. Following synthesis, we acetylated the peptide—acetic anhydride (3 eq) and diisopropylethylamine (4.5 eq) in dimethylformamide(2.63 ml) for 20 min—and then cleaved it from the resin with concomitant removal of side-chain-protecting groups by treatment with a cleavage mixture (10 ml) consisting of TFA (95%), triisopropylsilane (2.5%) and H₂O (2.5%) for 4 h at room temperature.

Suspended resin was removed by filtration, and the peptide was precipitated using three rounds of crashing in ice-cold diethyl ether, vortexing and centrifuging. The pellet was then dissolved in 1:1MeCN/H₂O and freeze-dried. Purification was performed by reverse phase high process liquid chromatography (RP-HPLC) using a Phenomenex Jupiter Proteo (C18) reverse-phase column (4 µm, 90 Å, 10 mm inner diameter × 250 mm long). Eluents used were as follows: 0.1% TFA in H₂O (a) and 0.1% TFA in ACN (b).

Peptides were eluted by applying a linear gradient (at 3.5 ml/min) of 5–95% B over 40 – 70 min. Fractions collected were examined by electrospray MS, and those found to contain exclusively the desired product were pooled and lyophilised. Analysis of the purified final product by RP-HPLC indicated a purity of >95%.

2.1.1: *b-f* Lactamisation

The introduction of lactam bridges to the peptides was achieved using Fmoc-Lys(Mtt)-OH and Fmoc-Asp(OtBu)-OPfp from Merck (Darmstadt, Germany) in the sequence at position ***b*** and ***f***, in lieu of Fmoc-Lys-OH and Fmoc –Asp(OtBu)-OH. Following aforementioned synthesis and acetylation steps, deprotection of the Asp(oPip) and Lys(Mtt) side chain protecting groups was achieved by repeated washing of the resin in dichloromethane, followed by repeated washes in dichloromethane (2% TFA), dichloromethane, and finally dimethylformamide. Resin was incubated for 6-8 hours at 55°C in 2-(1H-benzotriazole-1-yl)-1,1,3,3 tetramethyluronium hexafluorophosphate (1ml), diisopropylethylamine (1ml), and dimethylformamide (3ml). Resin was filtered and cleaved from the resin with concomitant removal of side-chain-protecting groups by treatment with a cleavage mixture (10 ml) consisting of TFA (95%), triisopropylsilane (2.5%) and H₂O (2.5%) for 4 h at room temperature. Following this, resins were removed by filtration, and the peptide was precipitated using three rounds of crashing in ice-cold diethyl ether, vortexing and centrifuging as before, with freeze-drying and purification following.

2.2 Circular Dichroism (CD)

Circular Dichroism is an absorption spectroscopy technique that utilises the fact that chiral molecules - when exposed to left and right circularly polarised light - display differential absorbance due to enantiomeric properties. As biological molecules (such as nucleic acids and peptides) are made up of a single optical isomer, they display this in the presence of differently polarised light. As CD is measured as a function of wavelength, the differences that spectrometers are able to detect are due to the presence of chromophores within these molecules.

Contributions to the profile of CD spectra differ depending on the part of the spectrum measured. Near-UV is considered to be between 250nm - 350nm and far-UV 190nm - 250nm. When considering proteins, the former monitors chromophores such as disulphide bonds and aromatic side-chains (such as those in tyrosine and phenylalanine). The far-UV region of the spectrum monitors the chromophore found in the carbonyl peptide bond (C=O). As a result, this region is sensitive to changes within the protein with respect to its folding state.

Peptides are made up of L-amino acids and the spectra produced when scanned through the range of 300 – 190 nm can be used to calculate their overall secondary structures, with distinctive profiles used to differentiate between the forms.

When considering α -helices, the spectra displays minima at both 208nm and 222nm. This, when combined with constant experimental technique (total protein concentration,

temperature, buffer solution, etc) can be used to calculate the level of α -helicity displayed and the differences found between multiple peptides.

Circular Dichroism experiments were conducted using an Applied Photophysics Chirascan machine (Leatherhead, UK). Samples contained 150 μ M total protein (Pt) in filter-sterilised buffer (10mM Potassium Phosphate and 100mM KF at pH 7). 200ul samples were contained in a Hellma Analytics (Mullheim, Germany) 10mM path length CD cuvette.

Raw CD units were converted to mean residue ellipticity (MRE) via the Equation 2:

$$\frac{\theta}{10 \times l \times [Pt] \times nr} \quad \text{Eq. 2}$$

Where θ = CD units, l = sample path length in cm (0.1), $[Pt]$ = peptide concentration in M (0.00015), and nr = average number of amino acid residues in the peptides.

All spectral data was converted to fractional helicity (fH) values according to Equation 3:

$$fH = \frac{\theta_{222} - \theta_c}{\theta_{222\infty} - \theta_c}$$

$$\theta_c = 2220 - (53 \times T)$$

$$\theta_{222\infty} = (-44000 + (250 \times k)) \times (1 - \frac{k}{Nr})$$

Eq. 3

Where the wavelength-dependent constant $k = 2.4$ (at 222nm), Nr = number of residues, and T = temperature ($^{\circ}\text{C}$)

2.2.1 Scans

CD Spectra were scanned between 300nm and 190nm in 1nm increments at 0.5 sec per increment. Measurements were taken at 20 $^{\circ}\text{C}$, 0 $^{\circ}\text{C}$, and at 20 $^{\circ}\text{C}$ following thermal melting to demonstrate the reversibility of binding.

2.2.2: Thermal Scans

For thermal scans (see Chapter 5), spectra were scanned between 300nm and 190nm in 1nm increments at 0.5 sec per increment. These scans were taken in triplicate at 10 $^{\circ}\text{C}$ steps

from 20°C to 90°C. Following this, a final measurement was taken at 20°C. Each temperature point was held for 1 min to equilibrate the sample before scanning.

2.2.3: Thermal Melts

Thermal denaturation experiments were taken by monitoring the signal of the sample at 222nm to observe any measurable change in the α -helicity of the peptide. The temperature range for these melts was 0°C – 90°C. These measurements were taken at 1°C increments, with 0.5 min for the samples to equilibrate.

Melting profiles were converted to equilibrium denaturation curves fitted to a two-state model using Equation 4:

$$\Delta G = \Delta H - \left(\frac{T_A}{T_m}\right) \times (\Delta H + R \times T_m \times \ln(P_t)) + \Delta C_p \times \left(T_A - T_m - T_A \times \ln\left(\frac{T_A}{T_m}\right)\right) \quad \text{Eq. 4}$$

This is a modification of the Gibbs–Helmholtz equation, where ΔG is the change in Gibbs free energy, ΔH is the change in enthalpy, R is the ideal gas constant (8.314 J mol⁻¹ K⁻¹), T_A is the reference temperature, P_t is the total peptide concentration, and ΔC_p is the change in heat capacity. Using this, it is possible to derive the melting temperature (T_m) as the point where the complex is 50% unfolded.

2.3: Size Exclusion Chromatography (SEC)

Size Exclusion Chromatography is a chromatography method that separates macromolecules based on their differing abilities to penetrate small pores present in the stationary phase. This allows separation by size, as size is the determinant for the efficiency by which macromolecules penetrate these pores.

Regarding proteins, applications of SEC include the characterisation of conformation, aggregation, and oligomerisation of the macromolecules.

Experiments were performed on a GE Healthcare ÄKTAexplorer using a Superdex Peptide 10/300 GL column (Buckinghamshire, UK) at room temperature. 100 μ l a 150 μ M sample in filter-sterilised buffer (10mM Potassium Phosphate and 100mM KF at pH 7) was injected and eluted at a flow rate of 0.5 ml/min. Elution profiles were recorded via A_{280} .

CHAPTER 3 - SOFTWARE DEVELOPMENT

Programs developed in this thesis were written in Python Version 3.5.2. For early development, scripts were run as individual .py files. Scripts were later compiled into application files (.exe) via Pyinstaller Version 3.4. All were run on a personal machine running Windows 8.1 with 12GB RAM.

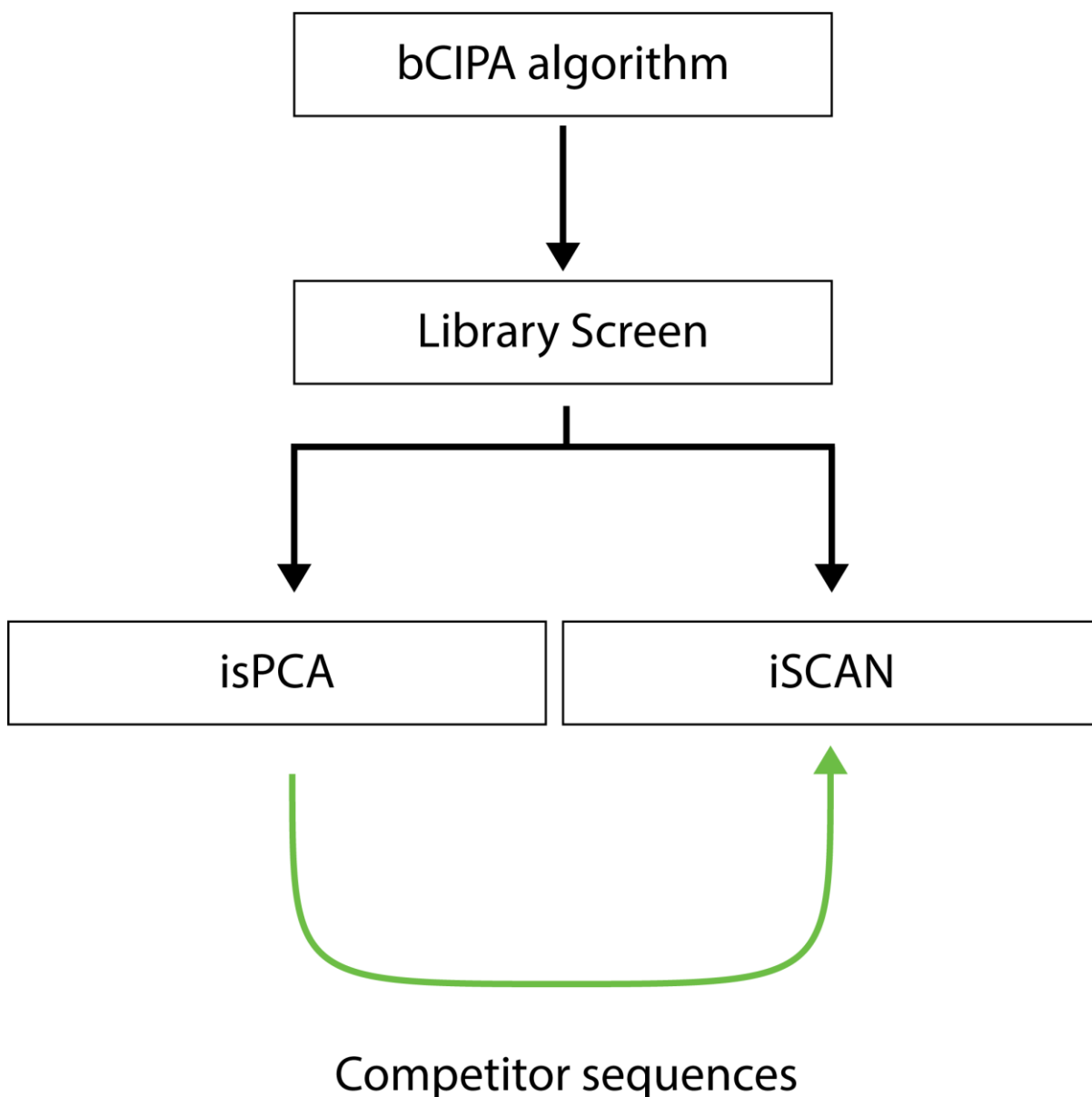


Figure 3.1. The overarching workflow for the *in silico* methods used in this thesis. The underlying architecture of the bCIPA algorithm is initially utilised to create a library screening module, allowing for library member sequences to be coupled with a single target to calculate a predicted T_m . This is further expanded upon in isPCA and iSCAN – synthetic analogues of PCA and CANDI – which incorporate comparative calculations with off-target states (with iSCAN including competitor sequences).

3.1 Library Screen

The library screening tool has been previously discussed (Crooks et al. 2011b; Lathbridge and Mason 2018) but, as shown in Figure 3.1, it is the most basic application of the bCIPA algorithm - which incorporates Helical Propensity (HP), Core (C) and Electrostatic interactions (ES) to provide a quantitative estimate of the interaction affinities in the form of a thermal melting temperature (T_m) as shown in Eqn.1:

Eq.1

$$T_m = (a \times HP) + (b \times C) + (c \times ES) + d$$

Within the library screening tool, the functions described below assigned scores to the peptide-peptide interactions based on this original bCIPA algorithm, including the coefficient values from least squares fit (Mason et al. 2006). The size of the coefficients ($a = 81.3256$, $b = -10.5716$, $c = -4.7771$ and $d = -29.1320$) acts as a modifier for the scale of the score – with the overall function for predicting the T_m defined as `calculate_tm`. The tool required an input from the user of the sequence that should be targeted but began previous to this by setting up the environment parameters for the software.

```
offset_dict = {
    'a':0,
    'b':6,
    'c':5,
    'd':4,
    'e':3,
    'f':2,
    'g':1,}

position = input("What is your heptad starting position (A-G)?: ")
rolling_offset = offset_dict[position.lower()]
```

The `rolling_offset` variable was set as the output from the `position` input string called into the `offset_dict` dictionary. This dictionary treated position **a** as the default starting point, with an assumption that any other starting position would represent elongation at the N-terminus into a previous heptad - hence the values increase from **g** - **b**.

```
raw_ESinteractionoffsets = [0, 0, 0, 0, -5, 0, +5]
raw_ESinteractionmodifiers = [0,0,0,0,1,0,1]
raw_Coreinteractionoffsets = [ 0, 0, 0, 0, 0, 0, 0]
raw_Coreinteractionmodifiers = [1, 0, 0, 1, 0, 0, 0]

ESinteractionoffsets = (np.roll(raw_ESinteractionoffsets,
rolling_offset))
ESinteractionmodifiers = (np.roll(raw_ESinteractionmodifiers,
rolling_offset))
Coreinteractionoffsets = (np.roll(raw_Coreinteractionoffsets,
rolling_offset))
Coreinteractionmodifiers = (np.roll(raw_Coreinteractionmodifiers,
rolling_offset))
```

The importance of the `rolling_offset` variable was highlighted in the modifier and offset values. The `modifiers` lists were binary. 1 denotes a position where the calculated value was included in further scoring, where 0 signifies an excluded position. The `offsets` were used to describe the location for the pairwise calculation to lookup. The positions contributing to the electrostatic interaction (**g** and **e**) had their interaction profiles (**g – e'+1**) denoted as -5 and +5 within the `raw_ESinteractionoffsets` list on the fifth and seventh positions respectively. These positions in `raw_ESinteractionmodifiers` were the only ones with values of 1. The hydrophobic core positions (**a** and **d**) had their inter-helical **a-a'** and **d-d'** interactions described in the `raw_Coreinteractionoffsets` list as 0 values, with the corresponding `raw_Coreinteractionmodifiers` positions all set to 0. This excludes the first and fourth position (representing the core positions), which were set to 1.

The role of `np.roll` was to shift the values in these 4 lists to reflect the starting position of the heptad defined in `position` and the resulting `rolling_offset` variable. As the command was called, `np.roll` shifted the values to the right, the `offset_dict` storage syntax was necessary.

With a starting position defined as **g**, the converted `ESinteractionoffsets` and `ESinteractionmodifiers` would – respectively – be:

```
[+5,0,0,0,0,-5,0]
[1,0,0,0,0,1,0]
```

As described in Figure 3.2, the calculation of the predicted T_m value was defined in the function `calculate_tm`. This required the incorporation of three scores – two of which were dependent on nested dictionary lookups (`CoreCalc` and `ESCalc`).

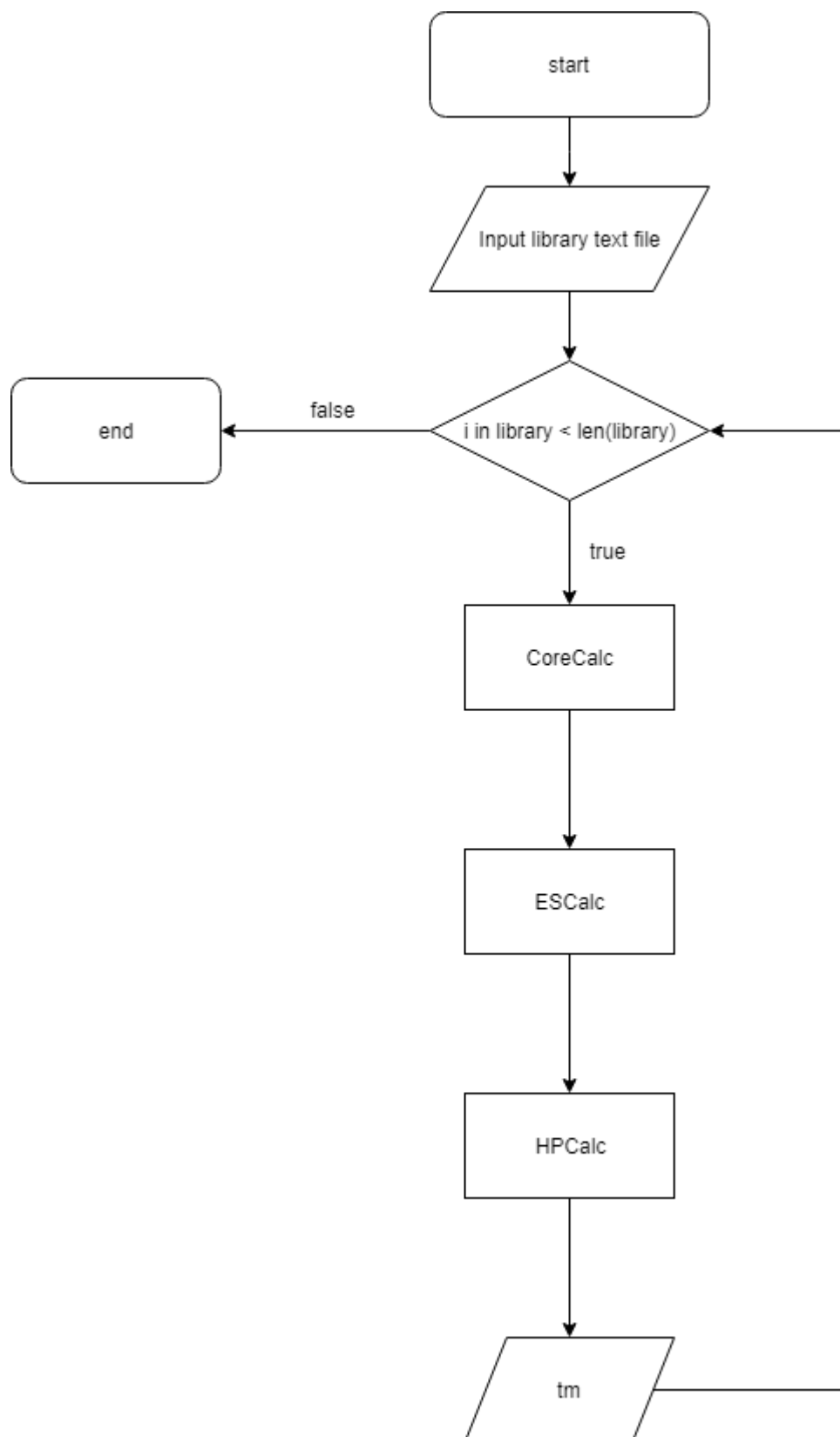


Figure 3.2. Overall flow diagram for the basic bCIPA library screen module. The `calculate_tm` function incorporated a loop of 3 main functions for calculating the contributions of **a/d** hydrophobic interactions, **e/g** electrostatic interactions, overall helical propensity (CoreCalc, ESCalc, and HPCalc). These scores were then used in the final calculation (t_m) for the prediction of the melting temperature of the interaction (T_m). This loop continued for the entire library, with the T_m and sequence data output as a `.txt` file after the final library member.

3.1.1 Core Contributions (CoreCalc)

To calculate the contribution of the core residues, the software had a function (`CoreCalc`) to incorporate the `Coreinteractionoffsets` and `Coreinteractionmodifiers` values assigned previously.

```
def CoreOffsets(peptideresidue):
    return Coreinteractionoffsets[peptideresidue]

def CoreInteractions(peptideresidue):
    return Coreinteractionmodifiers[peptideresidue]

def CoreCalc(peptide, target, core_table):
    numPlaces = min(len(peptide), len(target))
    peptideDif = len(peptide) - len(target)
    CoreScore = 0.0

    for i in range(numPlaces):
        residueA = peptide[i]
        heptad_pos = i % len(Coreinteractionoffsets)
        offset = CoreInteractions(heptad_pos)
        core_contribution_pos = i % len(Coreinteractionoffsets)

    core_interaction_mod=CoreInteractions(core_contribution_pos)

    residueB = target[i]

    CoreScore
    (core_table[residueA][residueB])*core_interaction_mod
    return CoreScore
```

A key part of this was the dictionary `core_table`, which contains 400 values defining the score assigned to a specific pairwise interaction from $\Delta\Delta\Delta G$ values calculated from the Vinson group thermodynamic scales (Acharya et al. 2006). These values were stored symmetrically, meaning that the lookup value for a Ile-Leu interaction was the same as that for a Leu-Ile interaction.

To begin with, the value of `CoreScore` (which was the total sum of the contributions of positions **a** and **g**) was set to 0.0 before the loop. The sequence fed in as `peptide` was from the text file containing the peptide library. During the loop, this sequence was accessed character by character. To keep the core interactions in frame (and avoid a lookup without a corresponding index), the loop was kept within `range(numPlaces)`, which was the length of the shortest peptide in the pair. Calling the `CoreInteractions` and `CoreOffsets` functions and defining the `core_contribution_pos` and `core_interaction_mod` variables, data was able to be fed into the final scoring calculation. `CoreScore` called a lookup of the `core_table` dictionary values for the string variables `residueA` and `residue`. This, combined with the binary `core_interaction_mod` value calculated the score for this pairwise residue interaction,

which was added to the overall `CoreScore` defined pre-loop. At the end of the function, this integer was the output variable and represented the scoring for the core contributions for the entire peptide-peptide interaction.

3.1.2 Electrostatic Contributions (ESCalc)

As for the core contributions, the calculation of the electrostatic contributions the software had a function (`ESCalc`) to utilise `ESinteractionoffsets` and `ESinteractionmodifiers`.

```
def ESResidueInteractions(peptideresidue):
    return ESinteractionoffsets[peptideresidue]

def ElectrostaticInteractions(peptideresidue):
    return ESinteractionmodifiers[peptideresidue]

def ESCalc(peptide, target, es_table):
    numPlaces = min(len(peptide), len(target))
    ESScore = 0.0
```

As before, the library `peptide` sequence was matched against the `target` sequence and the `modifier` and `offset` functions were called indirectly. This, coupled with the initialised `ESScore` variable (the equivalent of `CoreScore`), follows the same rules as the main function for core contribution calculation (`CoreCalc`). However, there were a different set of parameters incorporated in the `ESCalc` function to simulate the interaction profile of the residues on the **g** and **e** positions.

```
if (i + offset >= 0) and (i + offset < numPlaces):
    residueB = target[i+offset]
    ESScore+=(es_table[residueA][residueB])*(es_interaction_mod)
else:
    ESScore+= 0
return ESScore
```

This `if` statement found within the `ESCalc` function prevented out of frame errors during calculation of the `ESScore`. As the electrostatic positions are **e** and **g**, the `offset` value could be either -5 or +5. The first conditional (`i + offset >= 0`) prevented a residue at N-terminal **e** from looking back at a non-existent **g** (0 was the first character and `i+offset` with a non-existent **g** would have had a value less than 0). The second conditional (`i + offset < numPlaces`) prevented C-terminal **g** from attempting to pair with a non-existent **e** as it would index at a value greater than the length of the peptide sequence for interaction prediction (`numPlaces`).

As with core, the `ESScore` was updated with the value from the dictionary lookup – with the values contained within `es_table` scored from experimental $\Delta\Delta\Delta G$ data on the comparative contributions of residues at the electrostatic positions (Krylov et al. 1998b). This integer was stored and at the end of the function, it was returned.

3.1.3 Helical Propensity Contributions (HPCalc)

The contributions of each residue to overall helicity of the sequence were simpler within the calculations of the software.

```
def HPCalc (peptide, hp_table):
    numPlaces = len(peptide)
    HPScore = 0.0

    for i in range(numPlaces):
        residue = peptide[i]
        HPScore += hp_table[residue]

    return HPScore/len(peptide)
```

This function called a simple dictionary (`hp_table`) containing helical propensity values for each of the 20 natural amino acids (Williams et al. 1987). As before, `numPlaces` and `HPScore` were defined. The loop statement, for each residue in the sequence, called the value of the residue within the dictionary. This lookup value was then added to the `HPScore` and this value was returned as an average of the overall length of the peptide - allowing the helical contributions of sequences and interactions of different lengths to be comparative. This was the only function that was called twice, with the helical propensity of both the library member and the target sequence calculated independently and then combined to give the final HP value used in the overall T_m equation.

3.2 *In Silico* PCA (isPCA)

isPCA built on the original library screening framework by adding multiple simultaneous calculations (as opposed to a single loop of one calculation) to simulate the cellular environment in respect to the reformation of DHFR function. Importing the `calculate_tm` function incorporated the selection for the starting heptad position.

```
target = input()
library = input()
libname = library
library = GetLine(library)
lib = [i for i in (library)]
libsize = len(lib)
```

As before, the input for `target` was a single string whereas `library` was an input referring to a text file containing strings on each line. Similarly, the latter variable was updated to move the list into memory using the `GetLine` function. The `libname` variable was

assigned here as the name of the text file called in order to use later when formatting the data output. The library sequences were then read into another list (`lib`) which was then counted and stored in `libsize`.

```
best_PCA = []
bC_tm = []
bC_homo_tm = []
PCA_win = 0
maxdeltalist = []
mindeltalist = []
hetero_list = []

linescompleted = 0
lastpercentage = 0
start_clock = time.time()
basetime = start_clock
```

Following that, the program set up a series of empty lists which were populated as the prediction calculations initiated. The sequences which successfully met the parameters within the program (and were “isPCA successful”) were stored in `best_PCA`, the heterospecific T_m values (against the target) were stored in `bC_tm`, and their homodimeric T_m of these sequences were stored in `bC_homo_tm`.

```
for j, i in enumerate(lib):
    library_homo = calculate_tm(i,i)
    target_tm = calculate_tm(i,target)
    hetero_list.append(target_tm)
    better_than_homo = int(target_tm) - int(library_homo)
    better_than_target_homo = int(target_tm) - int(target_homo)
    deltalist = [better_than_homo,better_than_target_homo]

    if better_than_homo >= 0 and better_than_target_homo >=0:
        best_PCA.append(i)
        bC_tm.append(target_tm)
        bC_homo_tm.append(library_homo)
        maxdeltalist.append(max(deltalist))
        mindeltalist.append(min(deltalist))
    linescompleted += 1
    percentage_complete = int((linescompleted)/(libsize)*100)
```

An `enumerate` call over the library sequences (`i`) stored in `lib` was the main section of the isPCA prediction. To begin, the homodimeric T_m of the library sequence (library-library or L-L) and the heterospecific T_m of the sequence with the target sequence (library-target or L-T) were calculated and stored in the `library_homo` and `target_tm` variables, with the latter then being appended to the list `hetero_list` defined earlier. At this point, the two homospecific and one heterospecific T_m values required for the concept of PCA had been established, allowing for the selection of successful and unsuccessful sequences. In the `better_than_homo` calculation, the predicted T_m of the L-L interaction was subtracted

from that of the L-T interaction, with both values called as integers. Similarly, the `better_than_target_homo` calculation subtracted the predicted T_m of the T-T interaction from the predicted T_m of the L-T interaction. These two values were then stored in `deltalist`.

The following `if` statement sorts the successful peptides from the unsuccessful. This was because the former would need to have had “both `better_than_`” variables to be positive integers, showing that the L-T predicted T_m (desired complex) was higher than either of the homodimer complexes (off-targets). For the sequences that satisfy this statement, the sequence string was appended to `best_PCA` with the heterospecific and homodimeric predicted T_m values of the library members appended to `bC_tm` and `bC_homo_tm`. The homodimeric L-L and T-T values were sorted, with `maxdeltalist` appending the higher of the two integers and `mindeltalist` the lower of the two.

```
filedate = time.strftime("%d%m%Y-%H%M%S")
today = time.strftime("%b %d %Y")
path = ("PCA Output/" + (today))

if not os.path.exists(path):
    os.makedirs(path)

filename= ("PCA " + (project) + "_" + (filedate) + ".txt")
PCA_file = open(os.path.join(path,filename), "w+")
for num,peptide in enumerate(best_PCA):

    print("{} , {} , {} , {} , {}".format(peptide, (bC_tm[num]), (bC_homo_tm[num]
    )), (maxdeltalist[num]), (mindeltalist[num])), file=PCA_file)

PCA_file.close()
```

The isPCA data was output in a file created and opened by the `filename` and `PCA_file` variables. In `filename`, it called the `filedate` variable, which stores the date and time of the run and was put into a folder whose `path` was set by incorporating the date of the run. If the path doesn't exist, the command `os.makerdirs(path)` creates the sub-folder. The data for each sequence found in `best_PCA` were then stored into the output file. Using `num` as the index, each sequence was exported (`peptide`) with the sequence's specific values found in `bC_tm`, `bC_homo_tm`, `maxdeltalist` and `mindeltalist`. The output file was then closed.

```
print ("{} peptides retained and {} lost".format(PCAsize,PCAlast))
print ("\nPCA simulation was completed in %f seconds" % (time.time()
- start_time"))
```

Finally, the program printed the number of successful (PCAsize) and unsuccessful (PCAllost) sequences – along with the time data for the run.

3.3 *In Silico* CANDI (isCAN)

isCAN built on the framework set out in isPCA, with the addition of competitor sequences to provide additional off-targets in the form of library-competitor (L-C) and target-competitor (T-C) interactions. This can be described in an example AP-1 system with a targeted cJun sequence, screened against a Fos-based library (L) with natural Fos-family competitors (C).

```
phys_peptides = [cJun, JunB, JunD, cFos, FosB, Fra1, Fra2]
phys_names = ["cJun", "JunB", "JunD", "cFos", "FosB", "Fra1", "Fra2"]

jun_peptides = [cJun, JunB, JunD]
jun_names = ["cJun", "JunB", "JunD"]

fos_peptides = [cFos, FosB, Fra1, Fra2]
fos_names = ["cFos", "FosB", "Fra1", "Fra2"]

myc_peptides = [Myc, Max, Mnt, Mad]
myc_names = ["Myc", "Max", "Mnt", "Mad"]

myc_peptides = [Myc, Mnt, Mad]
myc_names = ["Myc", "Mnt", "Mad"]
```

isCAN stores a list of commonly used peptides in the Jun, Fos, and Myc family for use as competitor sequences. These were stored in paired lists, with `x_peptides` being the sequences and `x_names` being the name strings – used for the output of the interactions during and after the final calculations.

```
def choose_comp(competition_peptides):
    if competition_peptides == "1":
        phys_peptides = jun_peptides
        phys_names = jun_names
    elif competition_peptides == "2":
        phys_peptides = fos_peptides
        phys_names = fos_names
    elif competition_peptides == "3":
        phys_peptides = myc_peptides
        phys_names = myc_names
    elif competition_peptides == "4":
        phys_peptides = max_peptides
        phys_names = max_names
    elif competition_peptides == "5":
        phys_peptides = myc_peptides
        phys_names = myc_names
    else:
        phys_peptides = [cJun[comp_offset::], JunB[comp_offset::], JunD[comp_offset::], cFos[comp_offset::], FosB[comp_offset::], Fra1[comp_offset::], Fra2[comp_offset::]]
        phys_names = ["cJun", "JunB", "JunD", "cFos", "FosB", "Fra1", "Fra2"]
```

```
return (phys_peptides,phys_names)
```

The user choice for competitor sequences was an integer input which called a specific nested `if/elseif` statement. For an input of 1-5, the program chooses from the previously defined `x_peptides` and `x_names` lists, with all of the sequences used if the user does not state an input. These competitor strings were all modified by the `comp_offset` value to keep the positions aligned with the target and library sequences. As for `isPCA`, `isCAN` involved an `enumerate` call over the library of sequences, which then resulted in the L-T, T-T, and L-L predicted T_m values to be stored in the same output variables.

```
def negative_binding(peptide, offtargets):
    tm_list = []
    for off_target in offtargets:
        tm = calculate_tm(peptide, off_target)
        tm_list.append(tm)
    off_target_max = max(tm_list)
    off_target_min = min(tm_list)
    negative_output = [off_target_max, off_target_min]
    return negative_output
```

In addition to this, the `negative_binding` function was called to predict the T_m values for a sequence when in complex with each of the sequences that were chosen to be competitor off-targets. This resulted in a `negative_output` list which contains the highest (`off_target_max`) and lowest (`off_target_min`) values to be fed into the main loop.

```
for j, i in enumerate(lib):
    library_homo = calculate_tm(i,i)
    bind_the_comp = negative_binding(i,phys_peptides)
    compbindmax = bind_the_comp[0]
    compbindmin = bind_the_comp[1]
    target_tm = calculate_tm(i,target)
    hetero_list.append(target_tm)
    better_than_homo = int(target_tm) - int(library_homo)
    better_than_target_homo = int(target_tm) - int(target_homo)
    better_than_off_target = int(target_tm) - int(off_target_max)
    better_than_binding_off_target = int(target_tm) -
int(compbindmax)
    deltalist=
[better_than_homo,better_than_target_homo,better_than_off_target,b
etter_than_binding_off_target]

    if better_than_homo >= int(set_delta):
        if better_than_target_homo >= int(set_delta):
            if better_than_off_target >= int(set_delta):
                if better_than_binding_off_target >=
int(set_delta):
                    candi_win +=1
                    best_CANDI.append(i)
                    bC_tm.append(target_tm)
```

```

        bC_homo_tm.append(library_homo)
        bC_negative_tm.append(compbindmax)
        maxdeltalist.append(max(deltalist))
        mindeltalist.append(min(deltalist))
    else:
        fo_negative += 1
        compbindlist.append
        (better_than_binding_off_target)
        print ("Peptide {} negatively bound to {} @ {}
        degrees".format((j+1),list(phys_names[i] for i
        in ot_max_locations),(compbindmax)))
    else:
        fo_ot += 1
    else:
        fo_target_homo += 1
else:
    fo_lib_homo += 1

```

isCAN sequentially mimics the CANDI environment by incorporating calculations of off-target interactions for comparison with the desired complex (L-T). A series of nested `if/else` loops are applied and the outcome is dependent on the values calculated. This series of nested `if` functions were the parameters by which the successful peptides were selected, with repeat references to the user defined Δ value (`set_delta`). As shown in Figure 3.3, each stage compares the predicted T_m values of the various off-target states to that of the desired L-T interaction (`target_tm`). Unlike isPCA, which only required the predicted L-T T_m be higher than that of T-T or L-L homodimer T_m values, the presence of a competitor sequence in isCAN adds two extra potential off-target states to be considered (L-C and T-C). The successful peptides are output in a formatted file, with sequence strings and T_m values.

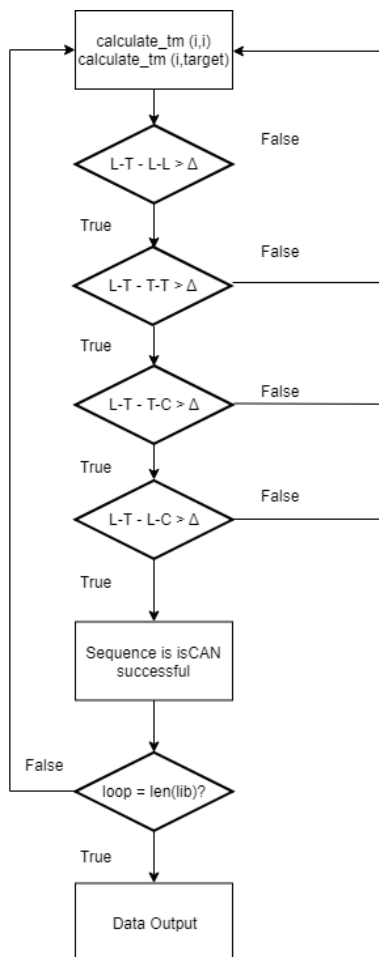


Figure 3.3 The main loop for the isCAN calculation through the library with parameters set by user. The loop initially called the `calculate_tm` function to predict the T_m value of both the library member homodimerically (i, i) and in complex with the target ($i, target$). This then loops through a series of true/false calculations which all call the values calculated by `calculate_tm` called for various interactions between the target(T), library member (L), and off target (C) and the Δ value set by the user. Δ refers the difference required between the T_m of the desired complex (L-T) and that of the individual off-target states (L-L, T-T, T-C, L-C). If, at any point, the calculated Δ value was lower than that of the defined Δ , a `false` state was output, the sequence was disregarded, and the loop re-initialises on the next library member. If all were `true` value, the sequence was considered “isCAN successful” and added to the output file. The loop continues until the final member of the library file had been through, at which point the data (including sequence and T_m integers) was output into a formatted `.csv` file.

3.4 Heptad Cassette Predictor (HCAP)

As HCAP was used on 7 residue sequences, a full length input sequence defined by the user and stored as `target` was length matched.

```

def heptadsplitter(target):
    n = 7
    heptad=[target[i:i+n] for i in range(0, len(target), n)]
    return heptad

```

Within the `heptadsplitter` function, `target` (a string variable) was segmented over the range of its individual sequence – `len(target)` – with every 7 characters being stored and returned in the `heptad` list. An example sequence `target` string of 35 characters (`RIARLEEKVKTLKAQNYELASTANMLREQVAQLKQ`) would result in a `heptad` output of `['RIARLEE', 'KVKTLKA', 'QNYELAS', 'TANMLRE', 'QVAQLKQ']`.

```
def plug_cassettes(target, library):
    tmlist = []
    for (pepnum, pephep) in enumerate(target):
        lst = []
        for (libnum, libhep) in enumerate(library):
            tm = calculate_tm(pephep, libhep)
            print("seqs      =      {}      (Target)\n      {}      (Library
Cassette)".format(pephep, libhep))
            print("{} , {} : {}".format(pepnum+1, libhep, tm))
            lst.append(tm)
        tmlist.append(lst)
    return tmlist

def homotm(library):
    tmlist = []
    for j, i in enumerate(library):
        tm = calculate_tm(i, i)
        print("{} , {} : {}".format(j+1, tm))
    return homotmlist
```

The inputs for the `plug_cassettes` function were the lists already defined in `target` and in `library`. To begin, the empty list `tmlist` was called to be filled. Following that was a pair of nested enumeration loops, with the top-level loop of `target`. For every instance of a target cassette string (`pephep`), an empty list (`lst`) was called and a sub-loop occurs over `library`. For HCAP to function, it called the `calculate_tm` function from the underlying library screening program, with `pephep` and `libhep` (sequences within the library) used for pairwise screening. For each iteration through the sub-loop, an integer `tm` value was added to `lst`. After each instance of the sub-loop, `lst` was appended to `tmlist` – creating a list of lists. If `library` contained 20 sequences and `target` contained 5, the returned `tmlist` would be a single list containing 5 lists and in each of those sub-lists, 20 T_m values would be contained. This was also done for the homodimeric cassettes, in which `library` was taken and each sequence was run against itself (`calculate_tm(i, i)`) – with the output (`homotmlist`) being a single list.

```

best_cassettes = []
for h,i in enumerate(tm1list):
    cassette_counter = 0
    for j, k in enumerate(i):
        print("Heptad {} vs Cassette {}: {}".format(h+1,j+1,k))
    tmhigh = i.index(max(i))
    tmhigh = tmhigh + 1
    best_cassettes.append(tmhigh)
print ("The best cassette combination is {}".format(best_cassettes))

```

In order to make sense of the data contained in `tm1list`, a final nested `enumerate` was called on this list of lists. For every instance of `i` (which refers to the sub-lists in which the predicted tm values were contained), the T_m values (`k`) were listed and printed. Following that, `tmhigh` takes the index of the highest value in each sub-list and appends it to the empty list `best_cassettes`. This was then increased by +1 in order to give the true index of the cassette (python initiates indexes from 0). At the end of the main `enumerate`, the program printed the output of this heterodimeric prediction to `output.txt`.

3.5 Library Comparison Tool

Library comparison allowed us to search for identical sequences between two separate libraries. These libraries were entered as text file called via two user inputted variables, `baselib` and `targetlib`.

```

baselib = input()
targetlib = input()
outputfile=open("{}-
{}.txt".format(baselib[:5],targetlib[:5]),"w+")
print("Peptides present in {} from the {} target library
\n".format(baselib,targetlib), file = outputfile)
targetlib = [i for i in GetLine(targetlib)]
targetsize = len(targetlib)
counter=0

with open(baselib,'rb',0) as file, \
    mmap.mmap(file.fileno(), 0, access=mmap.ACCESS_READ) as s:
    for num,peptide in enumerate(targetlib):
        if s.find(bytes((peptide), 'utf-8')) != -1:
            print("Found Target Peptide #{}".format(num+1))
            print("{}".format(peptide), file = outputfile)
            counter +=1
outputfile.close()
attrition = ((counter/targetsize)*100)
print ("Peptides found: {}".format(counter))
print ("This is {}% of the original target
library".format(attrition))

```

The program set up the output conditions, with `outputfile` opening a text file which was named after the first five characters of the two input libraries. This file was appended with a string showing the two different libraries used in the run. The list variable `targetlib` was

called as a list and was populated with the individual sequences from the text file input previously stored in `targetlib` (`[i for i in GetLine(targetlib)]`). Following this, `baselib` was accessed using `mmap` instead of the previously used `GetLine` function. This was because `mmap` supported files in memory-mapped objects – using the virtual memory of the operating system to access the data found in `baselib` rather than using separate calls to the system for each access, as well as the lack of necessity to copy the data between buffers. This was useful as `baselib` was more likely to be a larger file ($>10^6$), meaning that loading each sequence into memory was computationally inefficient when also loading `targetlib` as a second file for comparison. It allowed objects to constantly update and to be treated as mutable objects – with the ability to change and seek through sub-strings. With the `enumerate` through each sequence of `targetlib`, the element (`peptide`) was compared to the sequences (`s`) found in `baselib`. If a matching sequence was found, the index was printed non-pythonically (`num+1`) and the sequence itself (`peptide`) was appended to `outputfile`, which was then closed. The `attrition` variable was used to calculate the overall overlap between the two libraries and this was printed (along with `counter` – the integer count of sequences compared) to screen.

3.6 Quad Coil Interaction Prediction Algorithm (QCIPA)

For Chapter A1, QCIPA was developed to provide a tool with which an increasing number of heterospecific coiled coils could be grouped within a set for self-assembly. This tool was based on the framework defined by the library screening tool – with additional functions added for its purpose within a small library with minimal diversity (28 amino acids with Ile/Asn at position **a** and Lys/Glu at positions **e** and **g**) experimentally characterised previously (Crooks et al. 2016).

3.6.1 Anti-Parallel Coiled Coil Screening

One of the key functions within the software was the ability to remove peptides that were predicted to form anti-parallel homodimeric coiled coils. With the user input that these anti-parallel sequences should be removed, the following loop was called.

```
for (listnum,target) in enumerate(interactome_list):
    anti_parallel = str(reverse_word(str(peptide)))

anti_homo=((ESCalc((peptide),(anti_parallel),es_table)))
    if anti_homo == (ap_value):
        any_anti+= 1.0
        continue
    else:
        tm = calculate_tm(peptide,target)
        if tm>=desired_tm:
            cc_list.append([peptide,target])
```


The loop called the `ESCalc` function to score the electrostatic contribution (`anti_homo`) of a homodimeric anti-parallel interaction. If the user wishes to remove anti-parallel sequences, the variable `ap_value` was defined as the integer -12, representing the score for 8 instances of a Lys-Glu interaction (EK/KE was scored at -1.5) over 4 heptads. If the user does not, the variable was valued at -9999 (a value which could never come out in calculation). If `anti_homo` was calculated to be -12, the further prediction of T_m was not completed and the sequence was disregarded.

3.6.2 Progressive Coiled-Coil Set Expansion

Prediction of pair heterospecificity was revised from previous doubling methods (Crooks et al. 2016) to allow for a wider sweep of potential libraries fulfilling the criteria for predicted successful heterospecific sets. The incorporation of successive levels of sequences from a pool of heterospecific coiled-coils allowed for the software to avoid uninformative and simplistic levels of repetitively combining two libraries (for pairs -> quads -> octos).

```
pair_library = findpairs(cc_library)
tri_library = add_cc(pair_library,cc_library)
quad_library = add_cc(tri_library,cc_library)
quint_library = add_cc(quad_library,cc_library)
```

This series of additive functions resulted in expansion of the library until the parameters set by the user for stringency in the prediction of heterospecific pairs (such as maximum homodimeric T_m value) were no longer met. It was at this point that the final library was output as a text file with the other previous libraries which, at each step, were written to individual text files and purged from the software (to avoid iteratively larger and larger libraries from being held in memory).


CHAPTER 4 - COMPUTATIONAL COMPETITIVE AND NEGATIVE DESIGN TO DERIVE A SPECIFIC cJUN ANTAGONIST

Reproduced with permission from:

Lathbridge, A., and Mason, J. M. (2018) Computational Competitive and Negative Design to Derive a Specific cJun Antagonist. *Biochemistry* 57, 6108–6118

Copyright 2019 American Chemical Society.

Appendix 6B: Statement of Authorship

This declaration concerns the article entitled:			
Computational Competitive and Negative Design to Derive a Specific cJun Antagonist			
Publication status (tick one)			
Draft manuscript <input type="checkbox"/> Submitted <input type="checkbox"/> In review <input type="checkbox"/> Accepted <input type="checkbox"/> Published <input checked="" type="checkbox"/>			
Publication details (reference)	Lathbridge, A., and Mason, J. M. (2018) Computational Competitive and Negative Design to Derive a Specific cJun Antagonist. <i>Biochemistry</i> 57, 6108–6118		
Copyright status (tick the appropriate statement)			
I hold the copyright for this material <input type="checkbox"/> Copyright is retained by the publisher, but I have been given permission to replicate the material here <input checked="" type="checkbox"/>			
Candidate's contribution to the paper (provide details, and also indicate as a percentage)	<p>The candidate contributed to / considerably contributed to / <u>predominantly executed</u> the...</p> <p>Formulation of ideas:</p> <p>50% - Contributed <i>in-silico</i> ideas</p> <p>Design of methodology:</p> <p>80% - Developed software</p> <p>Experimental work:</p> <p>80% - Screening, characterisation, and data analysis</p> <p>Presentation of data in journal format:</p> <p>80% -Contributed to manuscript</p>		
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.		
Signed		Date	10/06/19

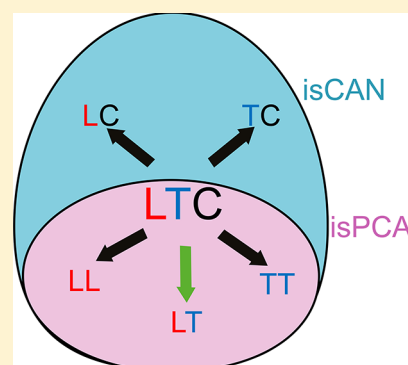
Computational Competitive and Negative Design To Derive a Specific cJun Antagonist

Alexander Lathbridge and Jody M. Mason*

Department of Biology & Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, U.K.

Supporting Information

ABSTRACT: Basic leucine zipper (bZIP) proteins reside at the end of cell-signaling cascades and function to modulate transcription of specific gene targets. bZIPs are recognized as important regulators of cellular processes such as cell growth, apoptosis, and cell differentiation. One such validated transcriptional regulator, activator protein-1, is typically comprised of heterodimers of Jun and Fos family members and is key in the progression and development of a number of different diseases. The best described component, cJun, is upregulated in a variety of diseases such as cancer, osteoporosis, and psoriasis. Toward our goal of inhibiting bZIP proteins implicated in disease pathways, we here describe the first use of a novel *in silico* peptide library screening platform that facilitates the derivation of sequences exhibiting a high affinity for cJun while disfavoring homodimer formation or formation of heterodimers with other closely related Fos sequences. In particular, using Fos as a template, we have computationally screened a peptide library of more than 60 million members and ranked hypothetical on/off target complexes according to predicted stability. This resulted in the identification of a sequence that bound cJun but displayed little homomeric stability or preference for cFos. The computationally selected sequence maintains an interaction stability similar to that of a previous experimentally derived cJun antagonist while providing much improved specificity. Our study provides new insight into the use of tandem *in silico* screening/*in vitro* validation and the ability to create a peptide that is capable of satisfying conflicting design requirements.



Coiled coils (CCs) are present in 3–5% of all amino acid structures and are highly versatile in the interactions they drive. They are characterized by a repeat of seven amino acids, the heptad repeat, with a preference for particular residue types at each position.¹ Despite the apparent straightforward link from their primary sequence to quaternary structure, CCs are highly specific in driving a wide variety of diverse protein–protein interactions, making them highly relevant systems in biotechnology and synthetic biology and as pharmacological targets. Parallel dimeric CC structures found within bZIP (basic leucine zipper) motifs are some of the simplest examples; they are comprised of two left-handed supercoiled α -helices that intertwine via a large interacting surface area.

Efforts to predict bZIP stability^{2–5} and, more recently, specificity^{2,4,6–8} are ongoing. The ability to predict the stability and specificity of peptides directly from the primary structure is a long-standing goal and is particularly important given the large number of human bZIPs. Current attempts to design peptides that specifically inhibit target leucine zipper interactions have taken an incremental library-based approach, with each new attempt improving our understanding of the factors that underpin the overall affinity. For example, *in vitro* assays of the binding affinity of 53 human bZIPs¹ showed there to be multiple interaction profiles, with specificity both within and between discrete bZIP families. This amounts to more than 1400 potential interactions, and selectivity within this set of bZIP interactions demonstrates how inherent sequence elements govern the selectivity of CC interactions. We have

previously utilized an intracellular library screening approach to derive specific antagonists of the oncogenic transcriptional regulator, activator protein-1 (AP-1).

Transcription factors represent compelling if difficult drug targets from conventional small molecule approaches. Their modulation can ensure that erroneous signals can be blocked at the transcriptional level, thus halting production of target genes implicated in disease, irrespective of the upstream signal imposed. Indeed, many oncogenic signal transduction cascades are known to upregulate transcription-factor activity, leading to gene expression changes that drive cell transformation^{9,10} and place bZIP families at center stage as therapeutic targets in cancer. Our previous efforts in this area have resulted in antagonism of AP-1 components by designing Jun- or Fos-based peptide libraries. This has been followed by their expression and screening inside living cells for an interaction with their target protein.^{3,11,12} In addition, we have experimented with methodologies in which competing off-target proteins are expressed in the assay during library selection. Both our conventional intracellular protein-fragment complementation assay (PCA) library screening approach^{3,11} and a target specificity-enhancing competitive and negative design initiative (CANDI)¹³ have resulted in a many PPI

Received: July 23, 2018

Revised: September 25, 2018

Published: September 26, 2018

inhibitors^{14,15} and CC-forming peptides in which the target is sequestered from binding to its natural partner. These assays have a significant advantage over *in vitro* approaches in enriching for target-specific sequences (relative to a broad range of other proteins expressed within the cell) and sequences that are structured, soluble, and nontoxic and resist protease breakdown. Moreover, the significant amount of data gained from these experiments, and our consequent improved understanding of the system, has facilitated the development of a series of tools that can work by predicting the affinities, and consequently the specificities, of CCs based only on input of the primary amino acid sequences.

The bZIP CC Prediction Algorithm (bCIPA)^{3,16} works by analyzing the helical propensities of component helices, together with the predicted contribution from electrostatic interactions and core residue interactions, to estimate the thermal denaturation temperature (T_m) of all hypothetical dimeric species within a defined system. Driven by coupling energies that describe a_i – a'_i hydrophobic interactions and g_i – e'_{i+1} electrostatic interactions as pairwise interactions measured by previous double-mutant analysis studies in CCs,^{17,18} bCIPA was derived to estimate the T_m of a given parallel dimeric CC using only the primary sequence and was shown to correctly predict 97% of all strong interactions and 95% of all non-interacting pairs using an independent data set of human bZIP proteins. As with previous prediction models, this approach allows prediction in a pairwise manner.

From previous work conducted to benchmark the accuracy of various prediction approaches using binding data from a FRET assay of interacting bZIP proteins,¹⁹ bCIPA has an accuracy similar to those of other purely data-driven models^{2,18} and had a better prediction accuracy for the 948 experimentally derived binding values than models that were driven solely by (or in combination with) CC structural prediction.^{20,21} Indeed, using the bCIPA engine, we have recently screened very large peptide interactomes to identify sets of ≤ 16 *de novo*-designed peptides that when combined are capable of forming specific CCs with their cognate partners.^{6,7}

Building on these previous findings, here we describe our efforts to take the approach much further by describing the first utilization of an *in silico* approximation to the PCA and CANDI-PCA approach, which we term *in silico* PCA (isPCA) and *in silico* CANDI (isCAN). The first approach allows the user (i) to define a target and (ii) to define every library member as a potential homodimeric off-target (Figure 1). The isCAN approach brings the additional capability of (iii) entering user-defined sequences that can interact with either the target or library member. Both isPCA and isCAN allow the user to create and *in silico* screen a library that is much more expansive than what can be accessed experimentally using either the complementary intracellular PCA or CANDI-PCA approach (i.e., $\sim 10^7$ for isCAN vs $\sim 10^5$ for PCA). The software then selects sequences on the basis of the greatest ΔT_m between all nondesired states and the desired target interaction, to give the highest predicted specificity. Here we describe the first implementation of the isCAN approach and provide an experimental validation of its use by computationally screening more than 60 million peptides to identify candidates that can bind to cJun specifically in the presence of cFos.

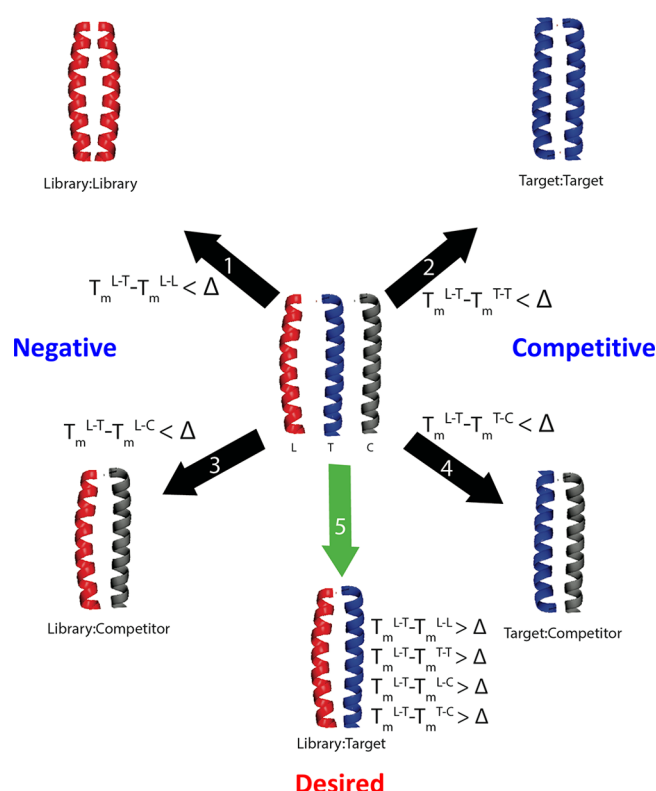


Figure 1. Overview of the CANDI protocol. Shown are the desired and numerous undesired states that can form upon combination of the library/target/competitor peptides. Complexes 1, 2 (negative), and 5 (desired) are found within PCA, with the competitor complexes (4 and 5) introduced in CANDI. Within isCAN, specificity is driven by the desired Δ value as specified by the user. The library member is successful only if it is able to form the desired complex with predicted T_m values greater than the Δ as specified by the user.

MATERIALS AND METHODS

In silico CANDI-PCA (isCAN) computationally screens a user-defined library against a given target. It identifies the highest predicted affinity binder to have the greatest difference between its target and off-target stability. This includes library homodimers as well as user-defined off-targets. isCAN utilizes the underlying bCIPA algorithm,^{2,3} which incorporates helical propensity (HP), core (C), and electrostatic interactions (ES) to provide a quantitative estimate of the interaction affinities in the form of a thermal melting temperature (T_m) as follows:

$$T_m = a \times \text{HP} + b \times \text{C} + c \times \text{ES} + d \quad (1)$$

The various functions within the algorithm assign scores to the peptide–peptide interactions. The size of the coefficients ($a = 81.3256$, $b = -10.5716$, $c = -4.7771$, and $d = -29.1320$) acts as a modifier for the scale of the score. For the calculation of HP, the average α -helical propensities⁴ of both peptides are calculated and totaled in eq 2:

$$f(\text{HP}) = \sum \text{HP}_a(l) + \sum \text{HP}_b(l) \quad (2)$$

The nature of the frame alignment that bCIPA employs ensures that, if the peptides are not of same length, the helical propensity is calculated to the length of the shorter peptide (l). For the calculation of the core interactions, only the residues within the hydrophobic interior (i.e., a or d positions) are considered such that the scoring mechanism is calculated accordingly:

$$f(\text{core}) = \sum \text{core}_{\text{res}} \times \gamma \quad (3)$$

The format of this function is such that the nonconsidered residues are calculated but are disregarded ($\gamma = 0$). Only for the a and d positions is the value of γ set to 1 (otherwise set to 0), ensuring that the core value in the final T_m calculation incorporates only these two heptad locations. As shown in eq 4, only positions e and g are considered in an g_i-e_{i+1} parameter when calculating the electrostatic parameters:

$$f(\text{ES}) = \sum \text{ES}_{\text{res}} \quad (4)$$

The program scans either peptide to calculate an electrostatic score for g_i-e_{i+1} and e_i-g_{i-1} . This ensures that, even in the case of different length peptides, all of the potential electrostatic interactions are taken into account in the final ES score without unnecessarily incorporating a substituent score more than once.

Although the bCIPA engine has been previously employed to provide an *in silico* interactome prediction algorithm for the derivation of heterospecific coiled coil sets,^{6,7} all previous implementations of bCIPA have been restricted to estimating the T_m for single pairs of peptides forming a CC. Considering the ability that the algorithm displayed in accurately distinguishing interacting from non-interacting CCs, the logical next step was to expand its remit to mirror the semirational design and screening approach used in an experimental setting. isCAN simulates the CANDI extension of the PCA screening technique and is a more advanced application of the bCIPA algorithm. The CANDI application of PCA involves the addition of competing peptides.¹³ If the target or the library member favors complexes with the competitor peptide, cell growth is either reduced or halted. Similarly, isCAN can consider multiple off-targets in addition to the target. To achieve this, in addition to built-in frame alignment and prediction functions, isCAN has a number of unique built-in check points. These make use of the individual predictions relating to the library (L), target (T), and competitor (C) peptides. Because of the optimization of core and electrostatic residues found in designed libraries, many peptides members are predicted to be more stable as homodimeric complexes than as heterodimers with the target. isCAN is split into two sections: the first set of calculations mirroring the PCA (isPCA section) and the second introducing the competitor peptides (isCAN). This stepwise calculation ensures that the processing time is not wasted on library members that are predicted to preferentially homodimerize or are unable to overcome the target homodimer (and are therefore not “PCA-successful”). A key concept in both is the predicted difference in T_m values (Δ). It is the key determinant behind the separation of successful and unsuccessful peptides in the library. User-defined, this value underlies all of the check points that the software considers. In particular, if $T_m^{L-T} - T_m^{L-L} > \Delta$ (i.e., the difference between the L–T desired heterodimer and the library homodimer is greater than the previously established desired Δ value), then the peptide is considered homodimerically successful and proceeds to the subsequent stages. Any peptides that meet both this and the target homodimer Δ ($T_m^{L-T} - T_m^{T-T} > \Delta$) are considered “PCA-successful” (i.e., complying with the scenario found during a PCA). The PCA-successful library members then have their desired state T_m (T_m^{L-T}) compared with CANDI-specific competitive library off-target T_m values (i.e., $T_m^{L-T} - T_m^{T-C} > \Delta$, and $T_m^{L-T} - T_m^{L-C} > \Delta$) and the “CANDI-successful” library members are

next exported for further analysis. Because of the multiple ways in which AP-1 may actively form, users can enter other Fos and Jun family member sequences to impose target specificity upon the screen. This addresses one of the key points of computationally aided peptide design with large families of peptides, avoiding interactions with other bZIPs that may be transcriptionally active and beneficial.

Calculating Off-Targets. In any simulated CANDI system, the interactions can be expressed as $2n + 3$, where n refers to the number of peptides introduced as competitive molecules, with only one desired (L–T) interaction. As an example, a cJun-targeting library would utilize four Fos family members (cFos, FosB, Fra1, and Fra2) as competitors. As previously mentioned, this would result in 10 off-targets for each peptide. As such, this would be 10 T_m values that the single desired L–T complex T_m must be able to overcome with $\Delta > 0$. To maximize the efficiency of the tool, the initial calculations made by the prediction software are of the heterospecific (L–T) complex. The output of this simple screen is used to partition the library by the desired T_m , irrespective of the ability to outcompete predicted off-target complexes. These partitions are 10^6 in size and are used to break down the computationally expensive isCAN calculations into computationally less demanding processes; approximately 36000 calculations are processed in 1 min (with this value increasing over time as increasing amounts of calculation data are stored within the application).

Peptide Synthesis. Rink amide ChemMatrix resin was obtained from PCAS Biomatrix, Inc. (Saint-Jean-sur-Richelieu, QC). Fmoc L-amino acids and 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate or benzotriazol-1-yl-oxytrypyrrolidinophosphonium hexafluorophosphate were obtained from AGTC Bioproducts (Hessle, U.K.). All other reagents were of peptide synthesis grade and obtained from Thermo Fisher Scientific (Loughborough, U.K.).

Peptides were synthesized on a 0.1 mmol scale on a PCAS ChemMatrix Rink amide resin using a Liberty Blue microwave peptide synthesizer (CEM, Matthews, NC) employing Fmoc solid-phase techniques²² with repeated steps of coupling, deprotection, and washing (4 × 5 mL of dimethylformamide).

Coupling was performed as follows. Fmoc amino acid (5 equiv), 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate or benzotriazol-1-yl-oxytrypyrrolidinophosphonium hexafluorophosphate (4.5 equiv), and diisopropylethylamine (10 equiv) in dimethylformamide (5 mL) were subjected to 35 W microwave irradiation at 90 °C for 30 min.

Deprotection was performed as follows. Piperidine (20%) in dimethylformamide was subjected to 30 W microwave irradiation at 80 °C for 5 min. Following synthesis, we acetylated the peptide—acetic anhydride (3 equiv) and diisopropylethylamine (4.5 equiv) in dimethylformamide (2.63 mL) for 20 min—and then cleaved it from the resin with concomitant removal of side-chain-protecting groups by treatment with a cleavage mixture (10 mL) consisting of TFA (95%), triisopropylsilane (2.5%), and H₂O (2.5%) for 4 h at room temperature.

The suspended resin was removed by filtration, and the peptide was precipitated using three rounds of crashing in ice-cold diethyl ether, vortexing, and centrifuging. The pellet was then dissolved in a 1:1 MeCN/H₂O mixture and freeze-dried. Purification was achieved via reverse-phase high-performance liquid chromatography (RP-HPLC) using a Phenomenex Jupiter Proteo (C18) reverse-phase column (4 μm, 90 Å, 10

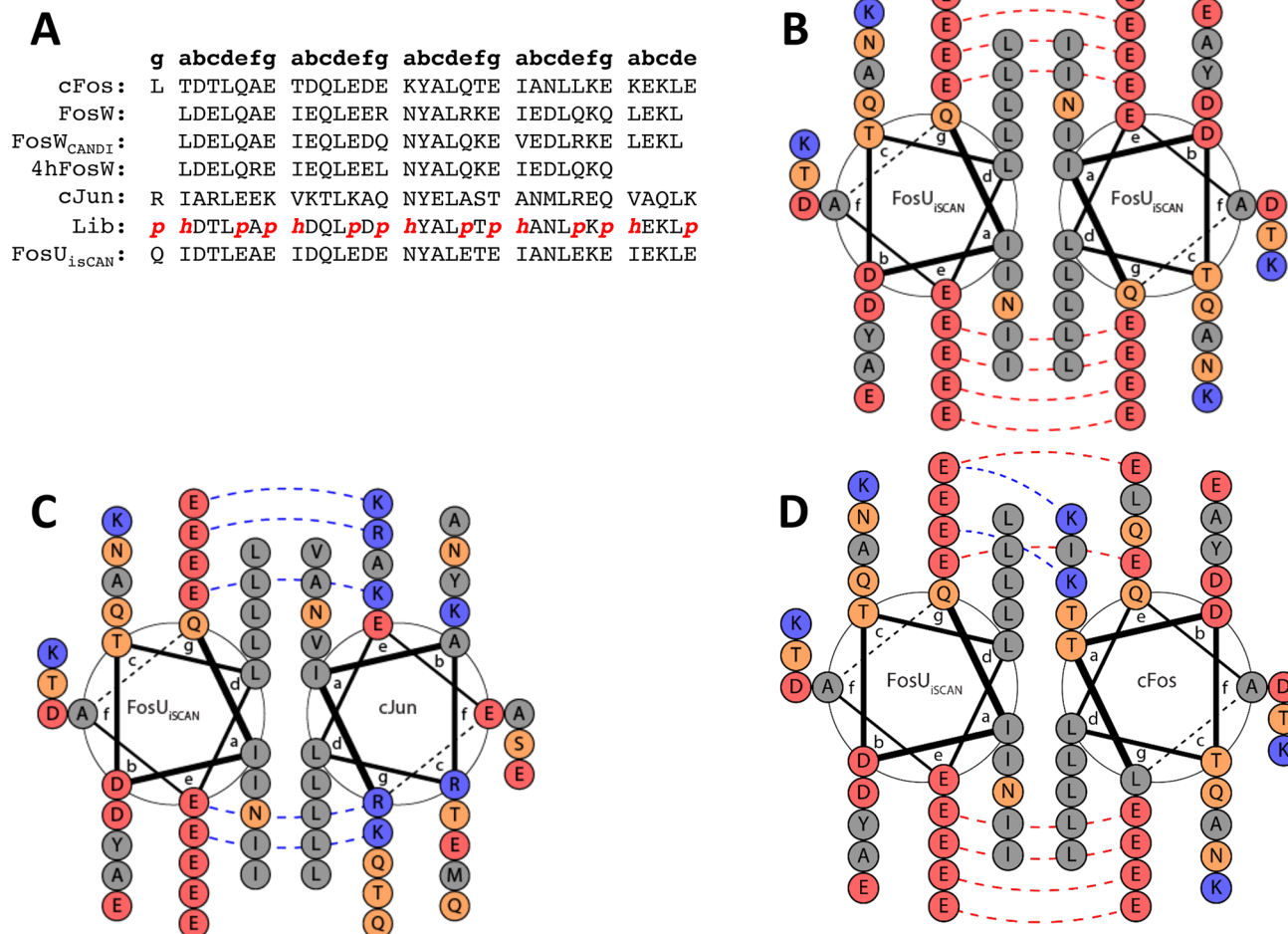


Figure 2. (A) Design of peptide inhibitor sequences. Peptide options are randomized around positions g and e (p) and position a (h). Compared to previous designs, the FosU_{isCAN} has had extensions added at N-terminal and C-terminal positions to add two extra residues for extra electrostatic interactions. (B–D) The helical wheel diagrams, generated with DrawCoil 1.0,⁵ display the residues present on the coiled coil from the position of the N-terminus to the C-terminus, looking down the axis of the α -helices. These diagrams illustrate the hydrophobic interface at the core position (a/d) and the charged residues present at the flanking position (e/g). The repulsive residues found at the electrostatic positions in off-target complexes (b and d) are selected. The helical wheel diagram of FosU_{isCAN}-cJun (C) demonstrates how FosU_{isCAN} has favorable electrostatic and core interactions to drive coiled coil formation.

mm inner diameter \times 250 mm length). The following eluents were used: 0.1% TFA in H_2O (a) and 0.1% TFA in ACN (b).

The peptide was eluted by applying a linear gradient (at 3.5 mL/min) of 5 to 95% B over 40 min. The fractions that were collected were examined by electrospray MS, and those found to contain exclusively the desired product were pooled and lyophilized. Analysis of the purified final product by RP-HPLC indicated a purity of >95%.

Circular dichroism (CD) was performed using an Applied Photophysics (Leatherhead, U.K.) Chirascan CD apparatus using a 200 μL sample in a CD cell with a 1 mm path length. Samples contained a 150 μM total peptide (Pt) concentration at an equimolar concentration for heterodimeric solutions (i.e., 75 μM per peptide) and suspended in 10 mM potassium phosphate and 100 mM potassium fluoride (pH 7) for 30 min prior to analysis. The CD spectra of the samples were scanned between 300 and 190 nm in 1 nm steps, averaging 0.5 s at each wavelength. Three scans at 20 $^{\circ}\text{C}$ were averaged to assess helical levels and the CC structure. Raw data (ellipticities) were collected and averaged, and data were converted to molar residue ellipticities (MREs).

Thermal denaturation experiments were performed at 150 μ M in a buffer of 10 mM potassium phosphate and 100 mM potassium fluoride (pH 7). The instrument that was used was an Applied Photophysics Chirscan circular dichroism device. For all thermal denaturation experiments, a stepping gradient was set from 0 to 90 $^{\circ}$ C in 1 $^{\circ}$ C increments (except for cFos-containing complexes, for which the process stopped at 50 $^{\circ}$ C). Each temperature point was held for 0.5 min to equilibrate the sample before scanning the ellipticity at 222 nm. Melting profiles were converted to equilibrium denaturation curves fitted to a two-state model, derived via modification of the Gibbs–Helmholtz equation to yield the melting temperature (T_m).²³

Size-exclusion chromatography experiments were performed at room temperature using a Superdex Peptide 10/300 GL column (GE Healthcare Life Sciences) by injecting 100 μ L of a 150 μ M Pt sample in 10 mM potassium phosphate and 100 mM potassium fluoride (pH 7) at a flow rate of 0.5 mL/min. Elution profiles were recorded via A_{280} .

Table 1. Top 10 Peptides Ranked by ΔT_m Predictions Calculated by isCAN Screening, Representing the Top 0.01% of isCAN-Successful Peptides^a

isCAN #	Peptide <i>p hDTLpAp hDQLpDp hYALpTp hANLpKp hEKLp</i>	Library-Target T_m (°C)	Library-Library T_m (°C)	ΔT_m
1	Q IDTLEAE IDQLEDK NYALKTE LANLEKE IEKLE	92.5	39.3	52.0
2	K IDTLEAE IDQLEDK NYALKTE IANLEKE IEKLE	94.2	42.7	51.5
3	Q IDTLEAE IDQLEDK NYALKTE IANLEKE IEKLE	91.9	38.1	51.0
4	K IDTLEAE IDQLEDK NYALKTE LANLEKE IEKLE	94.8	43.9	50.9
5	K IDTLQAE IDQLEDK NYALKTE IANLEKE IEKLE	91.1	41.2	49.8
6*	Q IDTLEAE IDQLEDE NYALETE IANLEKE IEKLE	91.3	41.6	49.7
7	K IDTLQAE IDQLEDK NYALKTE LANLEKE IEKLE	91.8	42.4	49.3
8	Q IDTLEAE IDQLEDE NYALETE LANLEKE IEKLE	91.9	42.8	49.1
9	K IDTLEAE IDQLEDK NYALKTE NANLEKE IEKLE	88.2	30.6	48.0
10	K IDTLKAE IDQLEDK NYALKTE LANLEKE IEKLQ	88.4	35.9	48.0

^aFosU_{isCAN}6 was selected for validation (* and named FosU_{isCAN}) because of the presence of Glu residues at positions g and e (bold). These were predicted to have maximal beneficial interactions with cJun and maximal repulsion with off-targets (i.e., in complex with cFos and as a homodimer). Sequences additionally contain N-cap (AS) and C-cap (GAP) motifs not depicted here.

RESULTS AND DISCUSSION

cFos-Based Library Generation. We previously utilized a number of *in vitro*^{12,16} and *in cellulo*^{5,11} peptide library screening approaches to derive sequences capable of binding to the cJun target of AP-1. One of these efforts utilized a PCA approach with a library of 62000 members to result in a 37-residue cJun antagonist (FosW).³ Using FosW as a template for further library design, this was followed by a truncated variant, 4HFosW, that retained most of the interaction affinity.¹¹ More recently, we have taken this further by rationally designing helix-constrained variants to permit downsizing of the molecule while retaining binding affinity.²⁴ As a mechanism for further increasing target selectivity during selection, we have also expressed off-target homologous sequences during PCA selection. The CANDI-PCA approach works by maximizing the difference in the free energy of binding between the target and off-target complexes by removing nonselective library members.¹³

Here we present a powerful new approach, based on the bCIPA engine, to facilitate the *in silico* derivation of specific peptide antagonists. As an example, we have derived a 39-residue antagonist that is specific for the cJun target. The sequence incorporates two additional residues over earlier designs, one g one e residue located at the N-terminus and C-terminus, respectively (Figure 2A). These permit four additional electrostatic interactions and in doing so provide a greater scope for stabilization/destabilization of target/off-target complexes to enhance interaction specificity. Using this extended cFos sequence as our design scaffold (Figure 2A), we have implemented an *in silico* approximation of the PCA and CANDI-PCA *in cellulo* screening systems to allow a rapid prediction of peptide sequences that display high target specificity. The tools described in this study are freely available (see the Supporting Information).

During library generation, each position was inspected and options were placed in the library sequence that corresponded to core hydrophobic and electrostatic positions within the

heptad repeats (a, e, and g). This resulted in an *in silico* library size of 60466176 peptides (a = ILVN, e = QEK, and g = QEK), with N included at all a positions to give rise to potential specificity driving N–N pairs with the a3 position on the target helix and to mitigate against the formation of higher-order oligomers.^{25–28} The predicted ΔT_m parameter (defined as the difference between the T_m of the desired dimer and the closest undesired dimer) was set to 20 °C during parameter initialization, because this was found to be the lowest value that resulted in a library that could be screened within a reasonable time frame while retaining a large number of peptides predicted to be successful, such that library diversity was retained. We followed only members of the top 10⁶ (partition 1) following these initial calculations, as this provided computational efficiency [reducing the calculation time from ~3 days to 5 h (see the Supporting Information)] while selecting library members of the highest predicted target affinity. This step reduced the expansivity of the search prior to entry into the more stringent isCAN step, which additionally considered members of the cFos family that naturally interact with cJun (i.e., cFos, FosB, Fra1, and Fra2) as explicit off-targets. As a competitive step in the initial isPCA, additional consideration of potential library members as homodimers and the stability of the cJun target as a homodimer were simulated (Figure 1). During this step, many peptides formed predicted homodimers or were not suitably more stable than the predicted T_m of the homodimeric cJun target complex and were consequently unable to overcome the stringent desired ΔT_m . Once the isPCA section was completed, the 60466176-member library was reduced to 73124 peptides, a reduction to 0.12% of the original library. The predicted ΔT_m values for the PCA complexes drove this reduction, i.e., the difference between library–cJun (L–T) and cJun–cJun (T–T) or library–library (L–L) interactions. Each successful peptide in the pool that satisfied the ΔT_m parameter set was permitted to proceed. These were described as sequences with predicted T_m values for undesired states (T_m^{L-L} and T_m^{T-T}) at least 20

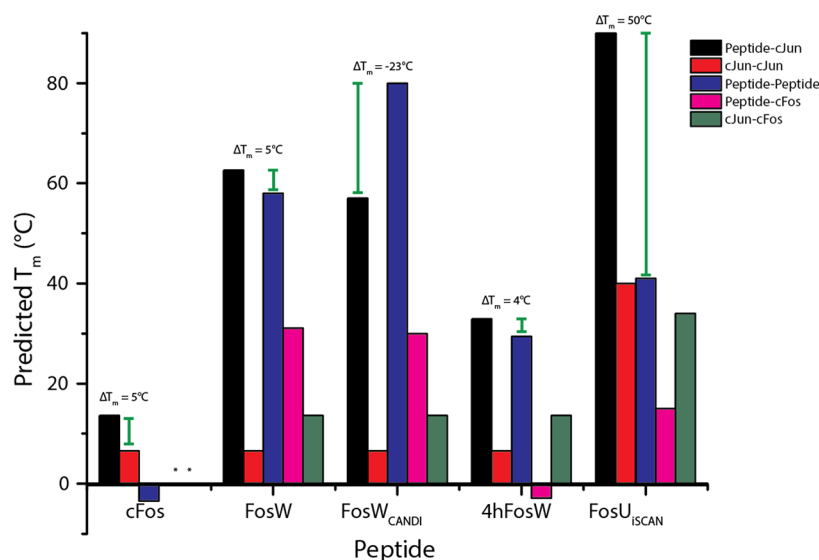


Figure 3. Predicted T_m values of the isCAN-selected peptide, FosU_{isCAN} is compared against previous Fos-based peptides targeting cJun using a cFos competitor (and cFos, not duplicating values*). All interactions were predicted using the same isCAN protocol. The ΔT_m values against the highest off-target (predicted library member homodimerization for all but cFos). FosU_{isCAN} is predicted to have a T_m of 91 °C with a ΔT_m of 50 °C; both of these values are the highest predicted for Fos-based peptides targeting cJun.

°C lower than the predicted library-cJun T_m (T_m^{L-T}). These “isPCA-successful” peptides were next entered into isCAN. This final step introduced simulated competitors [in this case, members of the Fos family that are known to form transcriptionally active complexes with cJun (cFos, FosB, Fra1, and Fra2)]. The isCAN step reduced the remaining library size further to 71667 peptides. The isCAN step was critical in removing 1457 members that were predicted to bind to one of the Fos off-target competitor peptides (T_m^{L-C}). These were again defined as those unable to overcome the required ΔT_m values among the library-cJun (T_m^{L-T}), target-competitor (T_m^{T-C}), and library-competitor (T_m^{L-C}) complexes. The average predicted ΔT_m for all L-C complexes (292496 interactions) was 11 °C.

Peptide Selection: FosU_{isCAN}. From the reduced size pool, peptides were finally ranked by the predicted ΔT_m value according to isCAN. This ensured that the peptide chosen for further study (FosU_{isCAN}) would exhibit both a high predicted T_m^{L-T} and a large ΔT_m (i.e., >20 °C) between this and the most stable of off-targets. As shown in Table 1, the peptide pool was reduced further to generate the top 10 sequences ranked by ΔT_m , which allowed for comparison of similarities and key differences between sequences.

These peptides represented the top 0.01% of all peptides to successfully emerge from isCAN. The final sequence, FosU_{isCAN}6 (termed FosU_{isCAN}), was selected for validation with the rationale that the high level of similarity between those of the 10 sequences and corresponding T_m values and ΔT_m values from nearest stability off-targets. The selected sequence was chosen on the basis of “charge blocks”, blocks of basic or acidic side chains at e/g positions that contribute favorably to the overall ΔT_m ^{6,7} but for which the sequence context of the otherwise energetically equivalent residue contribution is not considered by the software (see also below). Additionally, the minimal difference between predicted T_m and ΔT_m values within sequences listed in Table 1 meant that the predictive power of the software should be able to be validated without using the top peptide. As shown in Figure 2B–D, this potential inhibitor was not expected to form

interactions with off-targets and to be able to outcompete all possible other complexes (satisfying the competitive and negative design requirements of the experiment). As shown in Figure 3, the predicted ΔT_m between the closest off-target (in this case, L–L) and the desired (L–T) complex is 50 °C. The predicted T_m for the FosU_{isCAN}-cJun interaction (91 °C) is far greater than that of the closest undesired interactions (40 °C for the cJun homodimer and 41 °C for the FosU_{isCAN} homodimer).

isCAN Prediction. The isCAN-selected sequence (FosU_{isCAN}, QIDTLEA EIDQLED ENYALET EIANLEK EIEKLE) was predicted to be structurally optimized for maximizing and minimizing desired and undesired interactions, as shown using helical wheel diagrams (Figure 2B–D). For the negative design in avoiding formation of the FosU_{isCAN} homodimer complex, the electrostatic interactions play a vital role in destabilization. This peptide resulted in the introduction of e/g charge blocks,^{6,7} which was previously shown to be important in driving intramolecular repulsion between neighboring electrostatic side chains. We previously found that such charge patterns further assist in concomitantly driving both favorable interactions between antagonist and target and repulsions between potential antagonist homodimers (see also below), resulting in favorable gains in the measured ΔT_m . The introduction of these sequence-specific changes into antagonists otherwise considered energetically equivalent by the bCIPA approach can provide important contributions; they provide both intramolecular and intermolecular electrostatic contributions to stability that can concomitantly stabilize the desired state while destabilizing the homodimer. This is because neighboring residues with the same charge act to enhance or diminish the predicted energetics of intermolecular e/g interactions. This means that for homodimers the intramolecular repulsions act to enhance the intermolecular repulsions, making the complex less stable than is predicted without considering e/g residue sequence context. Concomitantly, for the desired heterodimer, the intermolecular repulsions act to assist the intermolecular attractions to a greater extent than predicted without e/g

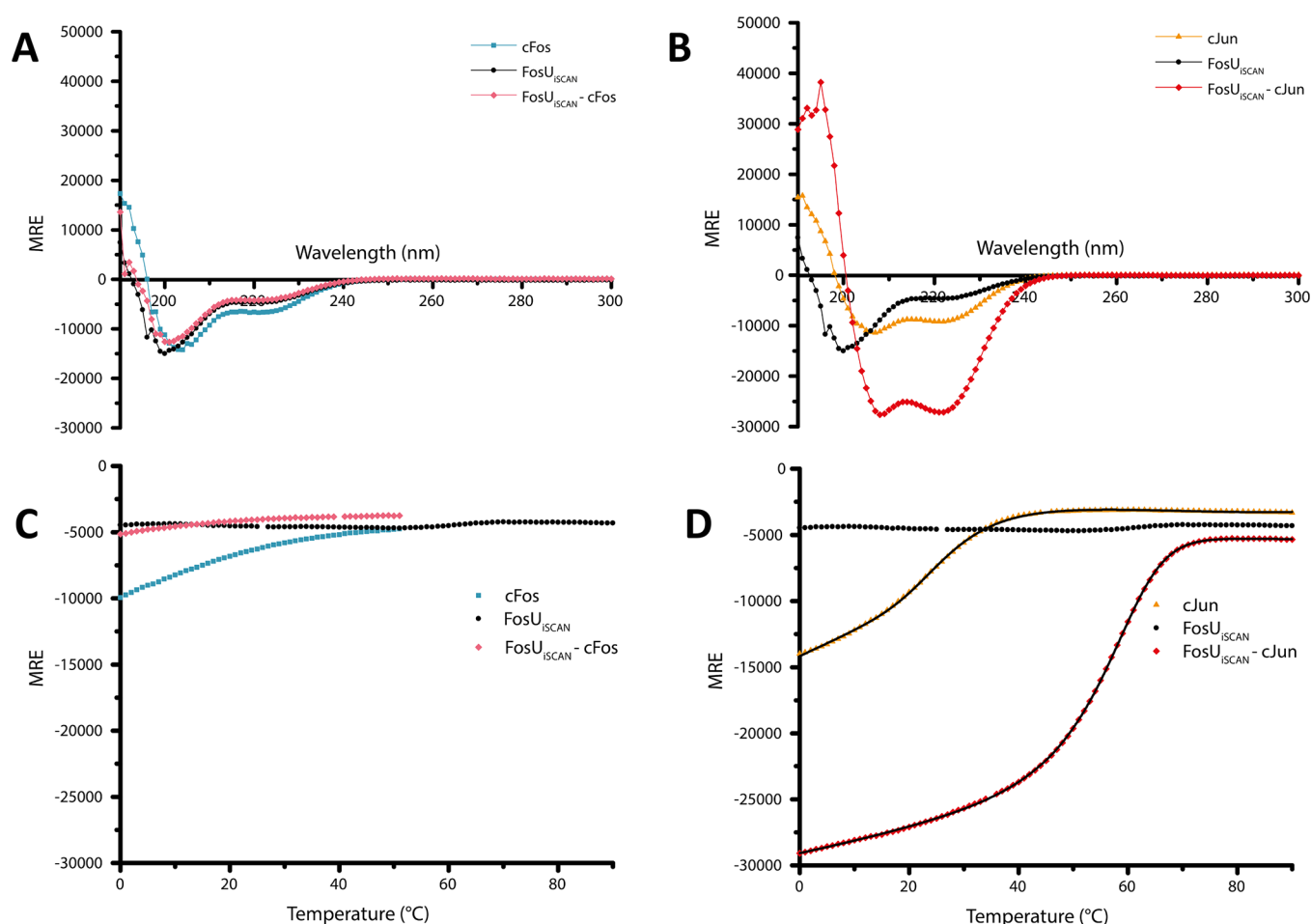


Figure 4. CD spectra and thermal denaturation data. Shown are data for the selected inhibitor peptide with (A and C) cFos and (B and D) target cJun. Spectra were measured at 20 °C at a total peptide concentration of 150 μ M and presented as mean residue ellipticity (MRE). The minima at 208 and 222 nm are indicative of a helical structure, with the 222 nm/208 nm ratio of the inhibitor with the cJun target showing more structure (222 nm/208 nm = 0.98) than the undesired complex with cFos (222 nm/208 nm = 0.56). This suggests that the inhibitor preferentially heterodimerizes with cJun. Thermal denaturation profiles of homodimeric peptides and FosU_{isCAN} with heterodimers (C and D) were taken using 1 °C increments and tracking the 222 nm signal at 150 μ M. FosU_{isCAN} shows an increase in the transition midpoint when in complex with cJun (D), demonstrating a T_m of 57 °C compared to the off-target state with cFos (C) that was unable to provide a measured T_m . All experiments were performed in 10 mM potassium phosphate and 100 mM potassium fluoride (pH 7). Where possible (D), data were fitted to the two-state model.

sequence context. The FosU_{isCAN} electrostatic interactions provide intermolecular charge blocks of four or five residues at g/e positions, which serves to add additional destabilization to the FosU_{isCAN} homodimer while adding additional stability to the target-bound heterodimeric complex. This is due to the presence of a Glu residue at both g^2 and e^3 , a combination not found in any other peptide within the top 10 from which FosU_{isCAN} was selected. As shown in Table 1, many of those specific g and e positions were populated by Lys residues. For FosU_{isCAN}, of the 10 possible electrostatic interactions between the residues at positions e and g, 50% contain favorably charged profiles (i.e., negatively charged in FosU_{isCAN} interacting with positively charged residues in cJun) and 50% have non-optimal profiles (negatively charged vs neutral or hydrophobic). For example, e^2-g^1 , g^1-e^2 , and g^4-e^5 all contain electrostatically favorable salt bridge interactions between oppositely charged Glu and Lys. Similarly, e^1-g^0 and g^3-e^4 feature favorable Glu–Arg interactions. Non-optimal profiles are a result of selection against residues native to the e and g positions of cJun, where negatively charged Glu is facing Ala, Gln, Thr, and Gln (e^3 , g^2 , g^3 , and g^4). As

shown in Table 1, inclusion of Gln at position g^0 for FosU_{isCAN} is found in three other peptides within the top 10. As this position is facing a Glu at e^1 (or a Gln in all Fos family members considered as off-targets), six of the peptides within the top 10 favorably target this by selecting Lys. However, Lys would form favorable interactions within the homodimeric complex by forming a g^0-e^1 Lys–Glu interaction (as in six other peptides found within the top 10). This is important because the cJun target peptide contains an Arg at position g^0 , meaning that the software has to decide on a g^0 residue selection driven by the optimal interaction with this Arg while balancing potential off-target interactions and selecting the option that will overall contribute to the greatest ΔT_m . The bCIPA algorithm scores an Arg–Glu interaction more favorably (−2.0) than an Arg–Gln interaction (−1.5) or an Arg–Lys interaction (−0.5). Therefore, it is more locally beneficial to have the e^1 position filled by Glu. Thus, to avoid favorable interactions in the homodimeric complex, it is locally favored to populate g^0 with Gln rather than Lys/Glu. Although there is a high level of sequence similarity within it, there are no peptides within the top 10 that differ from FosU_{isCAN} at a

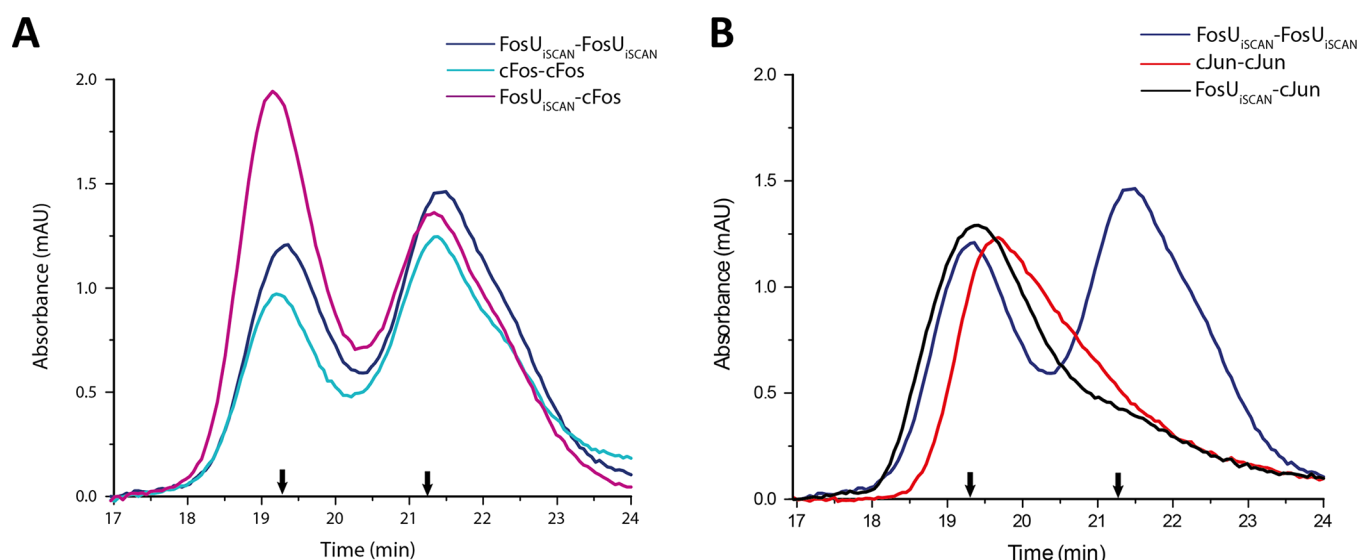


Figure 5. Size-exclusion chromatography experiments. Shown are SEC profiles for postmelt samples. (A) Peaks at approximately 19.3 and 21.3 min representing a mixture of dimer and monomer, respectively, in the FosU_{isCAN}-cFos mixture (dark blue). Component cFos and FosU_{isCAN} homodimer peaks show larger monomeric peaks than dimeric peaks. (B) The FosU_{isCAN}-cJun mixture generated a broad peak at approximately 19.3 min with the constituent cJun homodimer generating a peak at approximately 19.5 min, both indicating a dimer. Arrows show previously characterized monomer/dimer controls on 39-mer peptides.

single residue switch at g⁰ to give full Glu at all e/g positions. What is observed instead is the introduction of Lys within different heptads at different peptides. This is due to the aforementioned non-optimal residues on native cJun (e³, g², g³, and g⁴) with Lys-Glu and Lys-Gln interactions being scored favorably (both contribute -1.5).^{17,23} The Lys-Ala interaction has no associated electrostatic contribution value, and this is discussed below. This highlights part of the conflicting design requirements that isCAN attempts to address. At the core, cJun residues are optimized for hydrophobic interaction, with the a' position consisting of Ile, Val, Asn, Ala, and Val. FosU_{isCAN} takes advantage of this core arrangement with Ile at position a (with a³ as Asn to capitalize on the oligomer-limiting locus of the a³ N-N interaction²⁵). Across the top 10 peptides, the major difference is a⁴, where 40% of the peptides are Ile and 50% Leu (with one sequence selecting Asn) facing an Ala on a⁴. Both Ile and Leu contribute equally according to the algorithm (Ile-Ala/Leu-Ala = -0.5).

The stability of formation of the complex with the competitor molecule cFos is predicted to be low (Figure 2D), from both electrostatic and hydrophobic perspectives. The cFos core is poorly optimized for hydrophobic interaction (compared to transcriptionally functional cJun) because of the presence of multiple polar and charged Thr/Lys residues. Similarly, 60% of the cFos-FosU_{isCAN} electrostatic interactions are repulsive (Glu-Glu, +0.4 kcal/mol¹⁷). Moreover, the presence of Leu at e⁴ and g⁰ positions does not allow for further beneficial electrostatic interactions.

Circular Dichroism. An analysis of the global secondary structure content of homodimeric and heterodimeric systems was conducted at a total peptide concentration of 150 μ M to provide equimolar concentrations of each component helix for all dimeric systems. CD spectra showed FosU_{isCAN} to exist as a 15.4% weakly populated helical structure (Figure 4) with the 208 nm signal significantly exceeding that of 222 nm. Similarly, cFos (Figure 4A) and cJun (Figure 4B) existed as 20.6 and

27.5% helical structures with 222 nm/208 nm ratios of 0.60 and 0.82, respectively.²⁹

To analyze whether the selected peptide formed a complex with the cFos competitor sequence, a secondary structural scan of the FosU_{isCAN}-cFos complex was taken with CD (Figure 4A). As for component helices, this spectrum demonstrated the sample to lack both α -helical content (14.2%) and the double minima (222 nm/208 nm ratio = 0.56), indicating that the two component helices cannot associate to form a CC. Interestingly, both monomers involved in this heteromeric system displayed greater α -helical content when measured in isolation.

In contrast, the secondary structure content of the FosU_{isCAN}-cJun complex (Figure 4B) exhibited a much more intense signal with greater α -helical content (75.1%), almost 4 times stronger than the signal of the constituent peptides in isolation. In addition, the 222 nm/208 nm ratio was 0.98, providing further evidence for a significant increase in the helical stability of the sample. This demonstrates that the incubation of cJun with FosU_{isCAN} elicits a significant conformational change in the sample and provides compelling evidence for the formation of a CC.^{30,31}

Thermal Denaturation Profiles. Having observed a significant increase in the global secondary structure content of the cJun-FosU_{isCAN} sample, we next sought to quantify the stability of the complex by performing thermal denaturation experiments (Figure 4C,D and Table S2). In agreement with the spectra, this pattern of increased stability between undesired and desired complexes was also observed using thermal melts taken in 1 $^{\circ}$ C increments. FosU_{isCAN} in isolation did not form a CC complex; rather, only the upper baseline characteristic of the profile was observed (Figure 4C, black). This is in agreement with spectral data and is indicative of a weakly populated helix that does not self-associate. Further evidence for this is provided by size-exclusion chromatography (SEC), which demonstrates that the prominent species populated is monomeric (Figure 5A, blue). When FosU_{isCAN} is incubated with cFos (Figure 4C, pink), the thermal

denaturation signal is similar to that of the component peptides. In contrast, for homomeric cJun (Figure 4D, orange), a clear transition midpoint is visible (27 °C). Similarly, SEC experiments demonstrated cJun to exist as a dimer in solution (Figure 5B, red). However, when cJun was incubated with the FosU_{isCAN} antagonist peptide, the intensity of the helical signal increased significantly and led to an increased transition midpoint of 30 °C (Figure 4D, red), demonstrating an increase in thermal stability to 57 °C. This shift was further confirmed by SEC, demonstrating that the entire sample was in a dimeric state and that cJun had therefore paired with FosU_{isCAN} (Figure 5B, black). The inability of FosU_{isCAN} to form a stable homodimer in isolation is a considerable advantage, because it removes the homodimeric complex as a potential off-target. It is therefore able to form a stable CC only when combined with cJun.

The difference between the experimental thermal stability values and the values predicted by isCAN (Figure 3 vs Figure 6) is of interest. Via observation of the desired FosU_{isCAN}–cJun

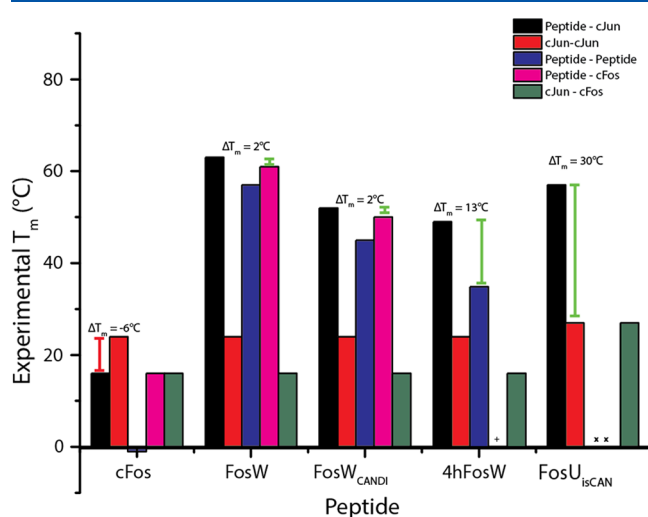


Figure 6. Measured T_m values of the isCAN-selected peptide. A comparison of measured FosU_{isCAN} with previously designed peptides FosW, FosW_{CANDI}, and 4hFosW. The T_m of the cJun–FosU_{isCAN} complex was measured as 57 °C with a ΔT_m of 30 °C from the cJun–cJun homodimer T_m of 27 °C. An x indicates neither FosU_{isCAN} nor its mixture with cFos was found to form dimers, with the thermal denaturation profile unable to provide a measured T_m ($T_m < 0$ °C). A plus sign indicates 4hFosW–cFos thermal denaturation data are missing (4hFosW homodimer data previously unpublished).

complex, there was a decrease of 34 °C between the predicted and experimental values. This is similar for 4hFosW and 4hFosW–cJun,²⁴ where the complex was measured to be 16–17 °C higher than bCIPA predicted. However, there was an observed decrease in the stability of the predicted extended cJun–cJun interaction that, combined with the off-target complexes, was not found to form a stable interaction. Overall, this means that the ΔT_m value has decreased from 50 °C when predicted to 30 °C when measured. Although this is a significant decrease, the measured value represents a larger difference between the desired state stability and nearest off-target stability than that for any previous inhibitor peptide we have developed. Previous work exploring the further biophysical characteristics of peptides with similar thermal stability through isothermal calorimetry²⁴ gives insight into the importance of this difference. The FosW–cJun complex (T_m =

63 °C) displayed a K_d value of 39 nM, whereas the 20HC–cJun complex (T_m = 33 °C) displayed a K_d of 15 μ M. Since this complex has a thermal stability similar to that of the FosU_{isCAN}–cJun interaction and the closest off-target, cJun–cJun, we can estimate that the 30 °C ΔT_m value denotes a sizable shift in the K_d from the range of nanomolar (desired state) to micromolar (off-target states).^{3,24}

bCIPA Screening. The discrepancies between T_m values predicted by isCAN and the measured values validated through CD suggest that predictions are overestimating the stability of some complexes, which are generally higher than those from experimentally measured thermal melts. A simple reason for this is the scope of the underlying bCIPA and how it was developed. Using the interaction profiles of 45 peptides (and tested on 59² interactions), it means that bCIPA has a wide scope for predicting peptide interactions.^{3,16} However, there are some instances in which the algorithm does not have the required data to estimate a contribution to binding affinity. For example, a Lys-Ala interaction is not estimated to make a contribution. Compared to a known and quantified interaction, it is an example of a non-optimal interaction. In comparison to interactions that are known to be nonfavorable (and thus positively scored), a value of 0 is considered more favorable, although it represents a lack of data. This may explain some of the discrepancies in predicted/actual T_m values observed, with the incorporation of energetically nonfavorable interactions. This is mitigated somewhat by the inclusion of the helical propensity values that each residue contributes, meaning that if the electrostatic contributions assumed within the algorithm are incorrect, there are other parameters that the algorithm uses to select residues.

From a software development perspective, the version used within this work stores all of the library sequences within the memory of the program as well as data generated by interaction calculations. Further development is ongoing, with the goal of minimizing the amount of data stored within the active program at any one point. It is hoped that this will limit the computational expense of this software and remove an obstacle in scaling up for high-performance systems. This would allow the isCAN approach to be used with larger systems and increased numbers of off-target peptides. Other *in silico* approaches with peptides have made use of other forms of searching within a large data set (including genetic algorithms³² and Monte Carlo methods³³). These methods of searching are typically coupled with molecular dynamics and docking simulations. If applied to a pairwise search with the appropriate methodology, this could represent a novel way to further screen for suitable peptides. Moreover, as we have previously demonstrated,⁶ bCIPA can be trained for specific bZIP subsets to increase its accuracy in such systems. Where knowledge of binding affinities is scarcer and more accuracy is required, an approach in which exploratory experiments are conducted and the data used to create the necessary bespoke training/test sets could be employed. This could allow the predictors within the algorithm to be adjusted accordingly for the bZIP profile.

Limitations of the Pairwise Approach. It has been suggested that the interaction of residues might not be limited to the pairwise model that bCIPA uses.¹⁹ Experiments that computationally derive additional scoring mechanisms from reported interaction affinities found that “triplet scoring”, the concept that the combinations of three residues among contact positions a, d, e, and g, could play a role in the prediction of

coiled coil interaction.¹⁹ Evidence that a combination of pairwise and triplet predictors increase accuracy provides further support for our “charge block” concept (blocks of same charge electrostatic residues at e and/or g positions).^{6,7} The charge block observations correlate to more nuanced, context-specific stabilization due to g/e residues interacting with a/d residues to modify the total interaction, and the underlying algorithm could be improved to reflect this. As described above, previous work on peptides of a particular profile⁶ has shown that training the algorithm on similar sequences has further optimized the weighting of the predictors to better predict T_m values. Because our studies have focused on Fos/Jun family bZIPs, a similar technique could be applied here. However, the lower T_m from predicted to measured values is consistent across the many FosU_{ISCAN} interactions studied in this system. This suggests that, although non-optimal for our elongated peptides, the software is still able to correctly predict T_m relationships. In comparison with previous work in this field, peptides generated solely through PCA and CANDI,^{3,13} this marks significant progress. As observed in Figure 6, although there is <10 °C between the measured T_m values of the cJun–4hFosW and cJun–FosU_{ISCAN} complexes, there is a measured increase of 17 °C in the ΔT_m , making FosU_{ISCAN} much more specific than 4hFosW for cJun relative to Fos. This value, indicative of the ability to design against negative and competitive states, while maximizing the desired state stability shows the real strength of the isCAN technique. Our aim was to create a competitive antagonist for cJun that, through high-throughput computational screening, would address conflicting design requirements between desired and undesired states. The increased ΔT_m of FosU_{ISCAN} relative to those of previous designs, coupled with a high thermal stability with cJun, supports our initial hypothesis that *in silico* screening of peptides to mimic and control the parameters of a PCA-CANDI can result in peptides that can selectively inhibit cJun without interacting with cFos or other off-target bZIPs. The off-targets are then free to form transcriptionally active components of AP-1. Future exploration into combining this approach with an *in cellulo* PCA-CANDI would be the next step in validating and potentially generating useful antagonists for future peptide therapies targeting not only AP-1 dysregulation but also any complex bZIP-mediated pathway in disease. This approach would provide a best-of-both combination of utilizing very large libraries, screening via a computational approach to enrich for predicted binders with the desired attributes of high affinity and selectively, and then finally experimentally screening the resulting reduced-size high-quality library that is accessible to intracellular selection systems.

In conclusion, our work provides a framework by which bZIPs can be modeled within a CANDI environment with accuracy to derive highly selective peptide sequences. Driving the approach with a solely computational and data-driven framework allows us to collect data about peptide–peptide interactions and specificity both within and between bZIP families. As more and more experimental data become readily available, this approach will become increasingly valuable in the design of specific peptides that can target key components within increasingly complex bZIP interactomes.

■ ASSOCIATED CONTENT

■ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.8b00782.

isCAN processing information, predicted melting temperatures (Table S1), experimental melting temperatures (Table S2), peptide sequences, and software download information (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: j.mason@bath.ac.uk. Telephone: +441225386867.

ORCID

Jody M. Mason: 0000-0002-4118-1958

Author Contributions

J.M.M. suggested and directed the research. A.L. developed the software, conducted the experiments, and synthesized, purified, and characterized the peptides. Both authors participated in experimental design, analysis of the data, and writing the paper.

Funding

J.M.M. is grateful to Cancer Research UK for a Career Establishment Award (A11738) and a Pioneer Award (A26941), to the BBSRC (BB/R017956/1), and to the EPSRC for an Overseas Travel Grant (EP/M001873/2). J.M.M. and A.L. thank the University of Bath for a Ph.D. Studentship.

Notes

The authors declare no competing financial interest.

■ REFERENCES

- (1) Mason, J. M., and Arndt, K. M. (2004) Coiled coil domains: Stability, specificity, and biological implications. *ChemBioChem* 5, 170–176.
- (2) Fong, J., Keating, A., and Singh, M. (2004) Predicting specificity in bZIP coiled-coil protein interactions. *Genome Biol.* 5, R11.
- (3) Mason, J. M., Schmitz, M. A., Muller, K. M., and Arndt, K. M. (2006) Semirational design of Jun-Fos coiled coils with increased affinity: Universal implications for leucine zipper prediction and design. *Proc. Natl. Acad. Sci. U. S. A.* 103, 8989–8994.
- (4) Grigoryan, G., Reinke, A. W., and Keating, A. E. (2009) Design of protein-interaction specificity gives selective bZIP-binding peptides. *Nature* 458, 859–864.
- (5) Grigoryan, G., and Keating, A. E. (2006) Structure-based prediction of bZIP partnering specificity. *J. Mol. Biol.* 355, 1125–1142.
- (6) Crooks, R. O., Lathbridge, A., Panek, A. S., and Mason, J. M. (2017) Computational Prediction and Design for Creating Iteratively Larger Heterospecific Coiled Coil Sets. *Biochemistry* 56, 1573–1584.
- (7) Crooks, R. O., Baxter, D., Panek, A. S., Lubben, A. T., and Mason, J. M. (2016) Deriving Heterospecific Self-Assembling Protein-Protein Interactions Using a Computational Interactome Screen. *J. Mol. Biol.* 428, 385–398.
- (8) Park, W. M., Bedewy, M., Berggren, K. K., and Keating, A. E. (2017) Modular assembly of a protein nanotriangle using orthogonally interacting coiled coils. *Sci. Rep.* 7, 10577.
- (9) Darnell, J. E. (2002) Transcription factors as targets for cancer therapy. *Nat. Rev. Cancer* 2, 740–749.
- (10) Rodríguez-Martínez, J. A., Reinke, A. W., Bhimsaria, D., Keating, A. E., and Ansari, A. Z. (2017) Combinatorial bZIP dimers display complex DNA-binding specificity landscapes. *eLife* 6, No. e19272, DOI: 10.7554/eLife.19272.
- (11) Crooks, R. O., Rao, T., and Mason, J. M. (2011) Truncation, randomization, and selection: Generation of a reduced length c-jun

antagonist that retains high interaction stability. *J. Biol. Chem.* 286, 29470–29479.

(12) Baxter, D., Ullman, C. G., Frigotto, L., and Mason, J. M. (2017) Exploiting Overlapping Advantages of in Vitro and in Cellulo Selection Systems to Isolate a Novel High-Affinity cJun Antagonist. *ACS Chem. Biol.* 12, 2579–2588.

(13) Mason, J. M., Müller, K. M., and Arndt, K. M. (2007) Positive aspects of negative design: Simultaneous selection of specificity and interaction stability. *Biochemistry* 46, 4804–4814.

(14) Acerra, N., Kad, N. M., Cheruvara, H., and Mason, J. M. (2014) Intracellular selection of peptide inhibitors that target disulphide-bridged A β 42 oligomers. *Protein Sci.* 23, 1262–1274.

(15) Cheruvara, H., Allen-Baume, V. L., Kad, N. M., and Mason, J. M. (2015) Intracellular screening of a peptide library to derive a potent peptide inhibitor of α -synuclein aggregation. *J. Biol. Chem.* 290, 7426–7435.

(16) Hagemann, U. B., Mason, J. M., Müller, K. M., and Arndt, K. M. (2008) Selectional and Mutational Scope of Peptides Sequestering the Jun-Fos Coiled-Coil Domain. *J. Mol. Biol.* 381, 73–88.

(17) Krylov, D., Barchi, J., and Vinson, C. (1998) Inter-helical interactions in the leucine zipper coiled coil dimer: pH and salt dependence of coupling energy between charged amino acids. *J. Mol. Biol.* 279, 959–972.

(18) Acharya, A., Rishi, V., and Vinson, C. (2006) Stability of 100 homo and heterotypic coiled-coil a-a' pairs for ten amino acids (A, L, I, V, N, K, S, T, E, and R). *Biochemistry* 45, 11324–11332.

(19) Potapov, V., Kaplan, J. B., and Keating, A. E. (2015) Data-Driven Prediction and Design of bZIP Coiled-Coil Interactions. *PLoS Comput. Biol.* 11, No. e1004046.

(20) Yang, Y., and Zhou, Y. (2008) Specific interactions for ab initio folding of protein terminal regions with secondary structures. *Proteins: Struct., Funct., Genet.* 72, 793–803.

(21) Rohl, C. A., Strauss, C. E. M., Misura, K. M. S., and Baker, D. (2004) Protein Structure Prediction Using Rosetta. *Methods Enzymol.* 383, 66–93.

(22) Fields, G. B., and Noble, R. L. (1990) Solid phase peptide synthesis utilizing 9-fluorenylmethoxycarbonyl amino acids. *Int. J. Pept. Protein Res.* 35, 161–214.

(23) Mason, J. M., Hagemann, U. B., and Arndt, K. M. (2007) Improved stability of the Jun-Fos activator protein-1 coiled coil motif: A stopped-flow circular dichroism kinetic analysis. *J. Biol. Chem.* 282, 23015–23024.

(24) Baxter, D., Perry, S. R., Hill, T. A., Kok, W. M., Zaccari, N. R., Brady, R. L., Fairlie, D. P., and Mason, J. M. (2017) Downsizing Proto-oncogene cFos to Short Helix-Constrained Peptides That Bind Jun. *ACS Chem. Biol.* 12, 2051–2061.

(25) Fletcher, J. M., Bartlett, G. J., Boyle, A. L., Danon, J. J., Rush, L. E., Lupas, A. N., and Woolfson, D. N. (2017) N@ a and N@ d: Oligomer and Partner Specification by Asparagine in Coiled-Coil Interfaces. *ACS Chem. Biol.* 12, 528–538.

(26) Gurnon, D. G., Whitaker, J. A., and Oakley, M. G. (2003) Design and characterization of a homodimeric antiparallel coiled coil. *J. Am. Chem. Soc.* 125, 7518–7519.

(27) Oakley, M. G., and Kim, P. S. (1998) A buried polar interaction can direct the relative orientation of helices in a coiled coil. *Biochemistry* 37, 12603–12610.

(28) Thomas, F., Niitsu, A., Oregioni, A., Bartlett, G. J., and Woolfson, D. N. (2017) Conformational Dynamics of Asparagine at Coiled-Coil Interfaces. *Biochemistry* 56, 6544–6554.

(29) Monera, O. D., Zhou, N. E., Kay, C. M., and Hodges, R. S. (1993) Comparison of antiparallel and parallel two-stranded alpha-helical coiled-coils. Design, synthesis, and characterization. *J. Biol. Chem.* 268, 19218–27.

(30) Lau, S. Y., Taneja, A. K., and Hodges, R. S. (1984) Synthesis of a model protein of defined secondary and quaternary structure. Effect of chain length on the stabilization and formation of two-stranded alpha-helical coiled-coils. *J. Biol. Chem.* 259, 13253–13261.

(31) Zhou, N. E., Kay, C. M., and Hodges, R. S. (1992) Synthetic Model Proteins - Positional Effects of Interchain Hydrophobic

Interactions on Stability of 2-Stranded Alpha-Helical Coiled-Coils. *J. Biol. Chem.* 267, 2664–2670.

(32) Heurich, M., Altintas, Z., and Tothill, I. (2013) Computational Design of Peptide Ligands for Ochratoxin A. *Toxins* 5, 1202–1218.

(33) Russo, A., Scognamiglio, P. L., Hong Enriquez, R. P., Santambrogio, C., Grandori, R., Marasco, D., Giordano, A., Scoles, G., and Fortuna, S. (2015) In Silico Generation of Peptides by Replica Exchange Monte Carlo: Docking-Based Optimization of Maltose-Binding-Protein Ligands. *PLoS One* 10, e0133571.

Supporting Information

Computational Development

All software was run using Python on a 64-bit x64-based processor Windows machine with 12 GB of RAM. Although no in-depth benchmarking was run due to time constraints, a library of 10^6 peptides required approximately 6 hours to fully output a successful library using isCAN. This became computationally expensive due to incomplete code optimisation discussed in the main text.

Stability Prediction

The prediction using the bCIPA algorithm resulted in a stable FosU_{isCAN} – cJun complex predicted with high levels of thermal stability (91°C) but with lower stability in the off-target homodimeric (41°C) and competitive (15°C) states. The T_m is predicted to overcome both of the transcriptionally active complexes of homodimeric cJun (40°C) and cJun-cFos (34°C).

	cJun	cFos	FosU_{isCAN}
cJun	40°C	34°C	91°C
cFos		-6°C	15°C
FosU_{isCAN}			41°C

Table S1. Predicted melting temperatures (T_m) by isCAN screening, taking into account electrostatic interactions, hydrophobicity and helical propensity.

Stability Measurements

Experimental values determined through thermal stability studies showed that the FosU_{isCAN} – cJun complex displayed a T_m of 57°C but with off-target homodimeric and competitive states that were unable to be determined through fit¹. The measured stability of cFos consistent previous work using a similar length cFos².

	cJun	cFos	FosU_{isCAN}
cJun	27°C	26°C	57°C
cFos		N/A (< 0°C)	N/A (< 0°C)
FosU_{isCAN}			N/A (< 0°C)

Table S2. Measured T_m values as validated using Circular Dichroism thermal denaturation analysis. The 222 nm signal was continuously monitored from 0-90°C. (homodimeric FosU_{isCAN} and in complex with cFos were unable to be determined through fit). cFos value obtained through a combination of methods as described in the supporting text.

Peptide Sequences

All synthesised and characterised peptides were amidated and acetylated and contained N- and C-capping motifs (underlined) for improved helix stability and solubility. Other Fos family peptides were considered for *in silico* calculations.

cJun: ASRIARLEEKVKTLKAQNYELASTANMLREQVAQLKGAP

cFos: ASLTDTLQAETDQLEDEKYLQTEIANLLKEKEKLEGAP

FosU_{isCAN}: ASQIDTLEAEIDQLEDENYALETEIANLEKEIEKLEGAP

FosB: LTDRLQAETDQLEEEKYELESEIAELQKEKERLE

Fra1: LTDFLQAETDKLEDEKYGLQREIEELQKQKERLE

Fra2: LTEKLQAETEELEEEKYGLQKEIAELQKEKEKLE

Software

The software used in this study (Library Creation and isCAN) is now freely available for Windows at <http://people.bath.ac.uk/jm2219/biology/downloads.htm>

References

- (1) Mason, J. M., Hagemann, U. B., and Arndt, K. M. (2007) Improved stability of the Jun-Fos activator protein-1 coiled coil motif: A stopped-flow circular dichroism kinetic analysis. *J. Biol. Chem.* 282, 23015–23024.
- (2) Mason, J. M., Müller, K. M., and Arndt, K. M. (2007) Positive aspects of negative design: Simultaneous selection of specificity and interaction stability. *Biochemistry* 46, 4804–4814.

4.6 Concluding Remarks on Chapter 4

This *in-silico* screening method has been effective in multiple ways, most notably in addressing the conflicting design requirements found in the development of peptide antagonists for AP-1. Particularly, the use of the Δ parameter to engineer instability within the off-target states – resulting in the inability of the homodimer and heterodimer with cFos complexes to form complexes that were able to form coiled-coils similar to the heterodimer with cJun.

While this constitutes a success in creating a purely *in-silico* antagonist, there are a few limitations to consider. Primarily, the discrepancy between the T_m predictions and the experimental values require further interrogation. The success in addressing the conflicting design requirements observed (instability in the off-target complexes with stability in the desired complex) lacks insight without addressing the discrepancy. This raises the question of whether the selected sequence was truly one of the most stable or if sequences outside of the top selected would have been better choices.

Secondarily, it does not offer insight into whether or not this antagonist is able to function in more complex environments. To address both of these limitations, the first step would be the addition of a previously designed *in-cellulo* assay such as PCA.

This would serve multiple purposes. Firstly, it would allow for further validation of the approach which, up until now, has relied on an *in-silico* to *in-vitro* pipeline. If a library sequence which is seen within the top *in-silico* peptides is then selected through PCA, it would justify the rationale of the *in-silico* technique. A potential addition of a sequence to target cFos would assess the effectiveness of the *in-silico* optimisation for a cJun target, allowing for hopeful steps into the area of co-compatible antagonists targeting both cJun and cFos.

Additionally, the *in-cellulo* success would address the selection conundrum. Due to the method used for PCA, the degenerate library expressed would have enough diversity to scan a wide range of sequences – with higher confidence in the selected peptide's stability (if corroborated with *in-vitro* thermal melt data).


CHAPTER 5 - COUPLING COMPUTATIONAL AND INTRACELLULAR SCREENING AND SELECTION TOWARDS CO-COMPATIBLE cJUN AND cFOS ANTAGONISTS

Reproduced with permission from:

Lathbridge, A., Michalowska, A. S., and Mason, J. M. (2019) Coupling
Computational and Intracellular Screening and Selection Towards Co-compatible
cJun and cFos Antagonists. *Biochemistry*

Copyright 2019 American Chemical Society.

Appendix 6B: Statement of Authorship

This declaration concerns the article entitled:			
Coupling Computational and Intracellular Screening and Selection Towards Co-compatible cJun and cFos Antagonists			
Publication status (tick one)			
Draft manuscript <input type="checkbox"/> Submitted <input type="checkbox"/> In review <input type="checkbox"/> Accepted <input checked="" type="checkbox"/> Published <input type="checkbox"/>			
Publication details (reference)	Lathbridge, A., Michalowska, A. S., and Mason, J. M. (2019) Coupling Computational and Intracellular Screening and Selection Towards Co-compatible cJun and cFos Antagonists. (<i>Biochemistry</i>)		
Copyright status (tick the appropriate statement)			
I hold the copyright for this material <input type="checkbox"/> Copyright is retained by the publisher, but I have been given permission to replicate the material here <input checked="" type="checkbox"/>			
Candidate's contribution to the paper (provide details, and also indicate as a percentage)	<p>The candidate contributed to / considerably contributed to / <u>predominantly executed</u> the...</p> <p>Formulation of ideas:</p> <p>60% - Contributed <i>in-silico/in-vitro</i> ideas</p> <p>Design of methodology:</p> <p>70% - Developed software</p> <p>Experimental work:</p> <p>80% - Screening, <i>in vitro</i> characterisation, and data analysis</p> <p>Presentation of data in journal format:</p> <p>80% - Contributed to manuscript</p>		
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.		
Signed		Date	10/08/19

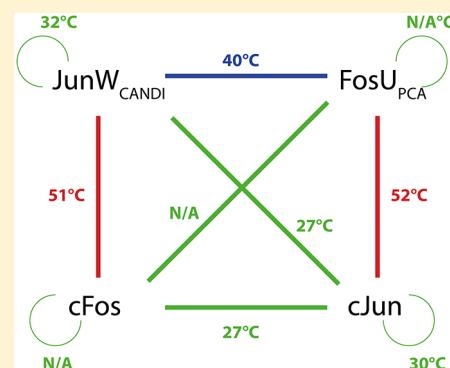
Coupling Computational and Intracellular Screening and Selection Toward Co-compatible cJun and cFos Antagonists

Alexander Lathbridge,[†] Anna S. Michalowska,^{†,‡} and Jody M. Mason^{*,†}

[†]Department of Biology & Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, United Kingdom

S Supporting Information

ABSTRACT: Basic leucine-zipper (bZIP) proteins represent difficult, yet compelling, oncogenic targets since numerous cell-signaling cascades converge upon them, where they function to modulate the transcription of specific gene targets. bZIPs are widely recognized as important regulators of cellular processes that include cell proliferation, apoptosis, and differentiation. Once such validated transcriptional regulator, activator protein-1, is typically composed of heterodimers of Fos and Jun family members, with cFos–cJun being the best described. It has been shown to be key in the progression and development of a number of different diseases. As a proof-of-principle for our approach, we describe the first use of a novel combined *in silico/in cellulo* peptide-library screening platform that facilitates the derivation of a sequence that displays high selectivity for cJun relative to cFos, while also avoiding homodimerization. In particular, >60 million peptides were computationally screened and all potential on/off targets ranked according to predicted stability, leading to a reduced size library that was further refined by intracellular selection. The derived sequence is predicted to have limited cross-talk with a second previously derived peptide antagonist that is selective for cFos in the presence of cJun. The study provides new insight into the use of multistate screening with the ability to combine computational and intracellular approaches in evolving multiple cocompatible peptides that are capable of satisfying conflicting design requirements.



Basic leucine-zipper (bZIP) transcription factors are a diverse family of DNA-binding proteins, typically consisting of a transactivation domain, a basic domain for binding to specific recognition elements within gene promoter regions, and a leucine-zipper (LZ) region that mediates bZIP dimerization required for activity. The transcription factor activator protein 1 (AP-1) can be composed of a wide range of dimeric complexes containing proteins that have physiologically varied roles linked to the control of gene regulation which impacts upon cellular proliferation and differentiation. Overexpression of AP-1 leads to oncogenic effects, such as dysregulation of the proliferation and differentiation of cells,^{1–3} making it and other related transcription factors compelling, if challenging, drug targets. Their dysregulation as part of oncogenic signaling pathways highlights the importance of specific targeting, while maintaining the ability to successfully modulate their activity in cellular transformation.^{4–7} The dimerization of bZIPs is mediated by the LZ domain, which contains a seven amino acid repeat (a heptad), which displays property-specific positions to drive the requisite interaction patterns.⁸ The coiled coil (CC) within bZIPs is comprised of two right-handed parallel α -helices that interact to form a left-handed supercoil structure, allowing key residues to periodically align in forming a heptad repeat every two turns of the helix. Despite the apparent simplicity of the CC, the relationship between its primary structure and the specificity found in its quaternary structure is not fully understood. Given the diversity and breadth of human bZIPs, there has been a

focus on engineering highly selective peptide-protein interactions.^{9–12} The creation of tools that can guide the user from the primary sequence to quaternary structure, along with quantitative information relating to interaction stability, is an ongoing effort.^{9,10,12,13} The role of AP-1 in cancer has made it a target of particular interest within therapeutic research, and multiple methods of inhibition have been explored.^{14–16} Previous work has explored designing antagonists based on Jun or Fos family peptides. A recent focus has been on the use of *in silico* approaches to simulate *in cellulo* screening techniques, with a view to predicting one peptide that satisfies the design requirements.¹⁷ Here we take this a step further by using the computational screens as a mechanism to reduce large libraries ($\sim 10^7$) to smaller higher quality libraries ($\sim 10^5$), which are then more readily accessible to intracellular screening approaches and predicted to contain many members with the desired properties, thus increasing the chances of success. This approach has made extensive use of the bZIP Coiled-Coil Prediction Algorithm (bCIPA)^{13,18} engine and focuses on creating *in silico* tools that mimic both the protein-fragment complementary assay (PCA)¹³ and competitive and negative design initiative (CANDI)¹¹ using bCIPA as an underlying algorithm to generate *in silico* PCA (isPCA)¹⁹ and

Received: July 24, 2019

Revised: December 5, 2019

Published: December 5, 2019

in silico CANDI (isCAN) equivalents.¹⁷ The use of CANDI as a framework for competitor mimicry promotes not only stability of a library-derived peptide in complex with its target but also the specificity required to avoid defined off-target states (i.e., more stable than predicted affinities of a target–target, or library member–library member interactions, or user-defined off-target–library-member interactions). Here, we describe a novel combination of both isPCA and isCAN from a large library followed by intracellular PCA on the refined library, with the aim of utilizing sequential screening capable of fulfilling the design constraints imposed by the need for specificity. In practice, this is 2-fold, the stability of the antagonist–target complex must be improved while also engineering a system in which the stability of the undesired complexes is decreased, thereby maximizing the difference between the multiple potential complexes. Additionally, we explore the capacity of the derived antagonist to complement a previously characterized antagonist that is selective for cFos in the presence of cJun (JunW_{CANDI}).¹¹ Importantly the design of the library described is such that it has the potential to target cJun while disfavoring binding to not only cFos but also potential library member homodimers. Moreover, the sequence and JunW_{CANDI} preferentially target their cognate binding partners over each other. Taken together, this raises the possibility to intentionally avoid cross-talk between either antagonist, library member–library member interactions (i.e., as homodimers), or the targets to which they bind. A key goal, therefore, is to offer bespoke cocompatible peptides with the potential for synergy in applying two peptides to simultaneously target a cJun–cFos AP-1 heterodimer remains.

Here we present our first efforts in this area, where we seek to address the following: (i) Does the combination of *in silico* and *in cellulo* screening result in a peptide able to successfully target cJun? (ii) Does the peptide chosen via a combined isCAN → PCA approach have increased target-specificity relative to previous *in cellulo* techniques? (iii) Can this peptide retain specificity in the presence of JunW_{CANDI}, an exemplar cFos antagonist that the library has not been screened against?

MATERIALS AND METHODS

***In Silico* PCA/CANDI (isPCA/isCAN).** The isPCA/isCAN computational screening techniques have been described elsewhere.¹⁷ Briefly, the approach mimics *in cellulo* PCA/CANDI screening techniques and simulates a defined library screened against a specific target sequence. Alongside the desired interaction with the target, the software considers homodimeric stability of both target and individual library members and, in the case of isCAN, a wide number of user-defined off-target stabilities. It identifies the predicted highest affinity binder (the desired threshold is set by the user), which also meets parameters for differences between predicted desired heterodimer T_m and that of all other complexes (also user-defined). Utilizing the underlying bCIPA algorithm, this pairwise analysis incorporates helical propensity (HP), core (C), and electrostatic interactions (ES) to provide quantitatively estimated values relating to the interaction affinities as thermal melting data (T_m).^{11,13,20,21}

Library Design and Cloning. Library design and cloning have been described elsewhere.²² Briefly, mega-primers were synthesized including relevant semirandomized codons for library residue options, and a fill-in reaction was performed, resulting in 140 bp double-stranded oligonucleotides. These were digested and cloned via NheI and Ascl sites into a pQE16

derivative (Qiagen) containing a G/S linker tagged to fragment 1 (p230d; Fos library; ampicillin resistance) or fragment 2 (p300d; cJun; chloramphenicol resistance) of murine dihydrofolate reductase (mDHFR), respectively. All proteins were under the control of a lac promoter, and expression was induced with isopropyl β -D-1-thiogalactopyranoside (IPTG). Library plasmids were transformed into BL21 cells (Stratagene) containing target plasmid and pREP4 (Qiagen; for lac repression; kanamycin resistance). To assess library quality, pools were sequenced collectively as well as single clones, and approximately equal distributions of varied amino acids were found. Pooled colonies were collected to exceed the library size 5–10-fold, to provide >95% library coverage.

Selection of Winner Peptides. PCA has been described in detail elsewhere.^{13,17,22} Briefly, target and library peptides are tagged at the genetic level to N- or C-terminal halves of the murine form of dihydrofolate reductase (mDHFR). Only two interacting helices will bring the two halves of the enzyme into close proximity, render the enzyme active, and result in colony formation on selective M9 minimal medium plates with 1 μ g/mL trimethoprim to selectively inhibit bacterial DHFR. Surviving colonies were pooled, grown, and serially diluted in liquid cultures under selective conditions (M9 minimal medium with 1 μ g/mL trimethoprim). The fastest growth, and hence the highest affinity interacting partners, was found to dominate the pool after 2 passages. Library pools, as well as colonies from taken from individual colonies isolated from competition selection pools, were sequenced to verify the arrival at one discrete sequence.

Peptide Synthesis and Purification. As described previously,^{17,23} Rink amide ChemMatrix resin was obtained from PCAS Biomatrix, Inc. (St.-Jean-sur-Richelieu, Canada); Fmoc L-amino acids and benzotriazol-1-yl-ox-ytripyrrolidinophosphonium hexafluorophosphate (PyBOP) were obtained from Merck; all other reagents were of peptide synthesis grade and obtained from VWR. Peptides were synthesized on a 0.1 mmol scale on a PCAS ChemMatrix Rink amide resin using a Liberty Blue microwave peptide synthesizer (CEM; Matthews, NC) employing Fmoc solid-phase techniques²⁴ with repeated steps of coupling, deprotection, and washing (4 \times 5 mL of dimethylformamide). Coupling was performed as follows: Fmoc amino acid (5 equiv), PyBOP (4.5 equiv), and diisopropylethylamine (DIPEA) (10 equiv) in dimethylformamide (5 mL) for 5 min with 35 W microwave irradiation at 90 °C. Deprotection was performed as follows: 20% piperidine in dimethylformamide for 5 min with 30 W microwave irradiation at 80 °C. Following synthesis, the peptide was acetylated using acetic anhydride (3 equiv) and DIPEA (4.5 equiv) in dimethylformamide (2.63 mL) for 20 min and then cleaved from the resin with concomitant removal of side-chain-protecting groups by treatment with a cleavage mixture (10 mL) consisting of TFA (95%), triisopropylsilane (2.5%), and H₂O (2.5%) for 4 h at room temperature.

The suspended resin was removed by filtration, and the peptide was precipitated using three rounds of precipitation in ice-cold diethyl ether, vortexing, and centrifuging. The pellet was then dissolved in 1:1 MeCN/H₂O and freeze-dried. Purification was performed by RP-HPLC using a Phenomenex Jupiter Proteo (C18) reverse-phase column (4 μ m, 90 Å, 10 mm inner diameter \times 250 mm long). Eluents used were as follows: 0.1% TFA in H₂O (a) and 0.1% TFA in ACN (b).

The peptide was eluted by applying a linear gradient (at 3.5 mL/min) of 5–95% B over 40 min. Fractions collected were

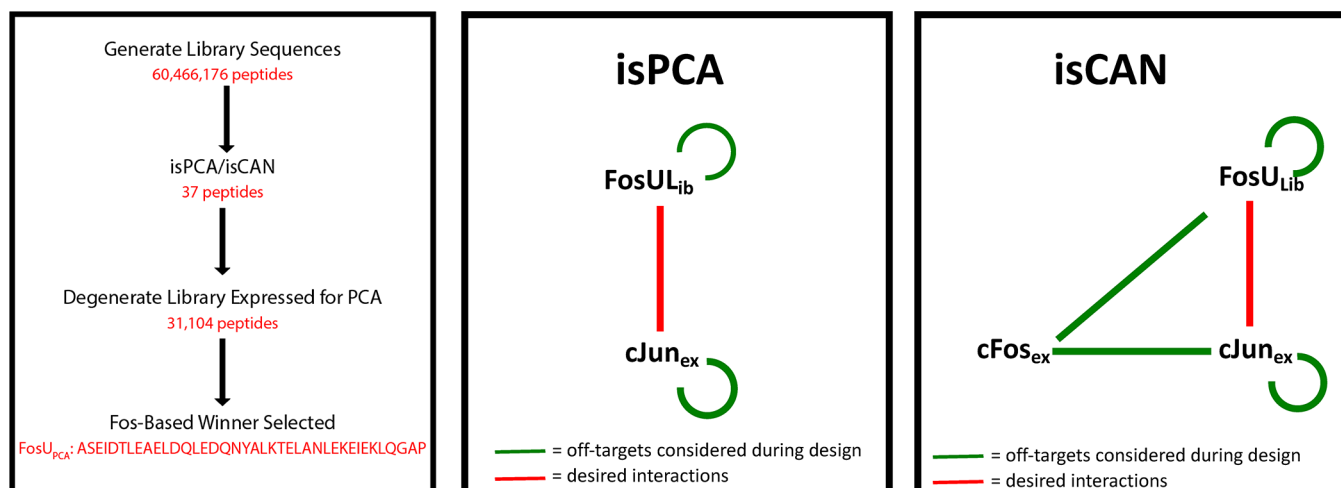


Figure 1. Overview of the combined *in silico/in cellulo* workflow. Left: A 60 466 176 member cFos-based library was generated and screened via isPCA and isCAN, using parameters and templates taken from previous work.¹⁷ The top 25 sequences from both were used to create a library (34 unique sequences due to overlap). This 31 104 (23 328 without the addition of 1 His at a^4) member library was expressed, and the winner sequence (FosU_{PCA}) was selected. Middle: Complexes that can be formed via the combination of library/target peptides (L/T). Negative and desired complexes are found within PCA, in considering the undesired homodimeric complexes (i.e., LL/TT). Right: isCAN incorporates the competitor complexes (i.e., the addition of LC/TC). Within isCAN, specificity is driven by the desired Δ value as specified by the user. The library member is only successful if it is able to form the desired complex with predicted T_m values greater than the Δ . In this Fos-based system, the competitor sequences included are from the FOS family (cFos, Fra1, Fra2, and FosB) for a total of 11 predictions for every library member simulation ($2n + 3$ where n = number of competitors).

examined by electrospray MS, and those found to contain the desired product exclusively were pooled and lyophilized. Analysis of the purified final product by RP-HPLC indicated a purity of >95%.

Circular Dichroism (CD). Analysis was performed using an Applied Photophysics (Leatherhead, U.K.) Chirascan CD apparatus using a 200 μ L sample in a CD cell with a 1 mm path length. Samples contained a 150 μ M total peptide (Pt) concentration at an equimolar concentration for heterodimeric solutions (i.e., 75 μ M per peptide) and suspended in 10 mM potassium phosphate and 100 mM potassium fluoride (pH 7) for 30 min prior to analysis. The CD spectra of the samples were scanned between 200 and 300 nm in 1 nm steps, averaging 0.5 s at each wavelength. Three scans at 20 °C were averaged to assess the overall helical content as well as the coiled-coil structure. Raw data (ellipticities) were collected and averaged, and data were converted to molar residue ellipticities (MREs).

Thermal Denaturation. Analysis was performed using an Applied Photophysics (Leatherhead, U.K.) Chirascan CD apparatus, recording the ellipticities of homotypic or heterotypic (1:1 stoichiometric mix) samples at a total peptide concentration (Pt) of 150 μ M in a buffer of 10 mM potassium phosphate and 100 mM potassium fluoride (pH 7). For all thermal denaturation experiments, a stepping gradient was set from 0 to 90 °C in 1 °C increments. Each temperature point was held for 30 s to equilibrate the sample to within 0.1 °C of the target temperature before measuring ellipticity at 222 nm. The resulting sigmoidal thermal denaturation profiles were fit to a two-state model, derived via modification of the Gibbs–Helmholtz equation to yield the melting T_m .²⁵

RESULTS AND DISCUSSION

In silico screening both without (isPCA) and with (isCAN) off-targets has been combined with PCA to derive a 39-mer peptide that is selective for cJun in the presence of cFos. From

thermal denaturation data coupled with dimer exchange experiments within the systems of the individual antagonists, the sequence is shown to be compatible with an existing sequence that is selective for cFos in the presence of cJun (37-mer JunW_{CANDL}).¹¹ Here we describe our approach toward co-compatible peptides that are capable of targeting specific components within the heterodimeric AP-1 complex with minimal cross-talk between partners and each other.

Library Creation and *In Silico* Screening. The *in silico* library used for this work was previously generated¹⁷ to create a highly expansive set that provided options capable of balancing both library diversity and simulation feasibility. In doing so, the ΔT_m parameter (the difference between the T_m of the desired complex and the closest undesired complex) was set to 20 °C during the initialization stage, leading to a library of 60 466 176 peptides. Briefly, this library contained semirandomized residues at the core and electrostatic positions, with LIVN options at a positions and QEK options at e and g positions. As previously described, the inclusion of Asn at all a positions was in order to mitigate against the formation of higher-order oligomeric states by driving the formation of Asn–Asn pairs with the a_3 position on the target helix.^{26,27} The peptides were next subjected to both isPCA and isCAN (Figure 1), resulting in libraries of 73 124 and 71 667 sequences, respectively. In order to create the library for PCA, the top 25 sequences (ranked according to a predicted ΔT_m of at least 20 °C and the value of the desired complex T_m) from isPCA and isCAN libraries were combined and used to define the library for PCA. Both the isPCA and the isCAN library contained sequences that represented 0.12% of the original 60 466 176 member library. Within this, the top 25 sequences represent 0.03% of these secondary libraries. The combination of the top 25 sequences from both resulted in 34 unique sequences due to an overlap of 16 sequences present in both. The library created from these 34 unique sequences resulted in a small high quality and PCA-accessible library of 23 328

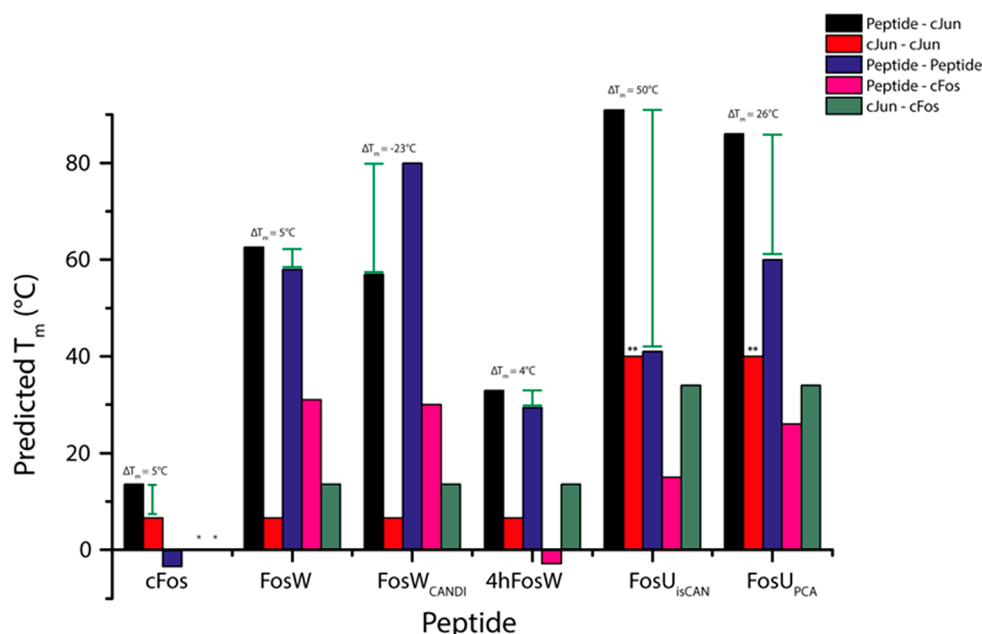


Figure 2. Predicted T_m values of FosU_{PCA} and related peptides. In comparison with Fos-based peptides targeting cJun with cFos as a competitor (* and cFos not duplicating values). All interactions have been predicted using the same iSCAN protocol. The ΔT_m values against the highest off-target (predicted to be library member homodimerization for all but cFos). FosU_{PCA} is predicted to have a T_m of 86 °C with a ΔT_m of 26 °C. ** For both FosU_{isCAN} and FosU_{PCA}, all sequences have been extended to 39 residues with N-terminal g and C-terminal e positions filled.

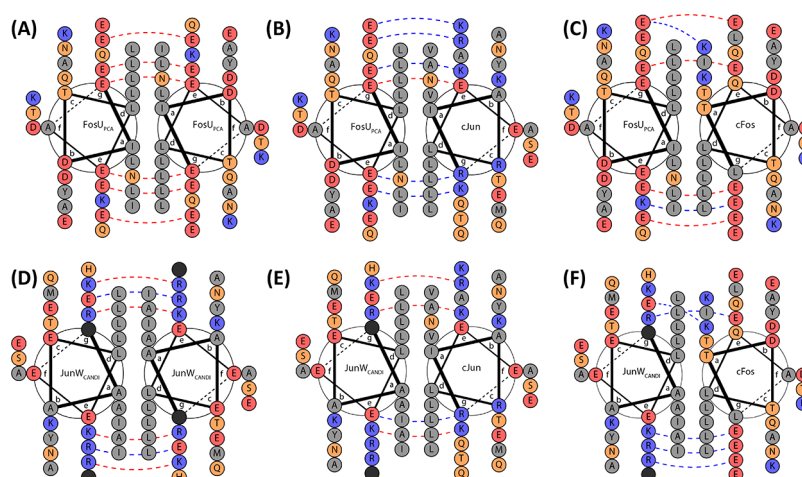


Figure 3. Helical wheels of FosU_{PCA} and JunW_{CANDI} interactions. The helical wheel diagram displays the residues present on the coiled coil from the position of the N-terminus to the C-terminus, looking down the axis of the α -helices. For both FosU_{PCA} (a–c) and JunW_{CANDI} (d–f), these diagrams illustrate the hydrophobic interface at the core position (a/d) and the charged residues present at the flanking position (e/g). The helical wheel of FosU_{PCA}–cJun (c) demonstrates how FosU_{PCA} contains residues that promote favorable electrostatic and core interactions to drive coiled-coil formation. JunW_{CANDI} in complex with cFos also shares this (f), with multiple attractive Glu–Arg and Glu–Lys electrostatic interactions.

sequences, which expanded further to 31 104 sequences with the addition of His at one core position (Figure S2). The inclusion of this residue was unavoidable when incorporating Asn (AAC) at a^4 with Ile (AUC) and Leu (CUC), requiring the use of two ambiguous nucleotides (MWC), with $M = A/C$ and $W = A/U$ (see Supporting Information)

PCA Selection. The stages of PCA demonstrate a stabilization of the semirandomized library through each passage, with the sequence identified from the final passage termed FosU_{PCA} (ASEIDTLEAELDQLEDQNYALKTE-LANLEKEIEKLQGAP). The details of the sequence highlight the result of the number of conflicting selection pressures enforced by the binding and growth assay, with progression

tracked in Figure S1. Prediction of the melting temperature (T_m) of FosU_{PCA} with its desired and off-targets (Figures 2 and 3) highlighted the diversity of its origins. As with previous work led by *in silico* design,¹⁷ the presence of a large difference (ΔT_m) between the desired complex (FosU_{PCA}–cJun) and the undesired complexes was enforced ($\Delta T_m = 26$ °C).

This predicted value was lower relative to that of previous work with this library due to the increased predicted T_m of homodimeric FosU_{PCA}–FosU_{PCA} ($T_m = 60$ °C) relative to that of the purely *in silico* derived FosU_{isCAN} ($T_m = 41$ °C). Although the FosU_{isCAN} was one of the top 25 sequences used as a basis for the library, it was not selected during PCA. The contrasting selections of *in silico* vs *in silico* → *in cellulo*

approaches are highlighted by the fact that the final selected sequence, FosU_{PCA}, is not observed in either of the top 25 sequences from isPCA or isCAN. Upon additional screening of the library, however, FosU_{PCA} was identified as being within the top 20% of all library predicted members within the PCA library (predicted ΔT_m and desired complex T_m).

Circular Dichroism Spectroscopy. The global secondary structures of FosU_{PCA} in isolation and in complex with cJun and cFos were analyzed to monitor for both helicity and interaction (Figure 4a/b). CD spectra showed that at 20 °C,

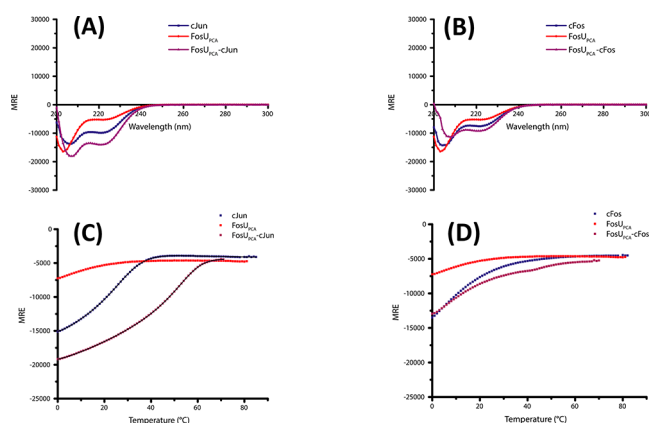


Figure 4. CD spectra and thermal denaturation data for FosU_{PCA} with cJun and cFos. Shown are data for FosU_{PCA} (red) with cJun (A and C, blue) and cFos (B and D, blue). Spectra were measured at 20 °C at a total peptide concentration of 150 μ M and presented as mean residue ellipticity (MRE). The minima at 208 and 222 nm are indicative of a helical structure when coupled with fractional helicity (fH), with the 222 nm/208 nm ratio of FosU_{PCA} showing less structure (222 nm/208 nm = 0.46 and fH = 17.0%) than cJun (222 nm/208 nm = 0.73, fH = 30.5%) and cFos (222/208 = 0.61, fH = 22.9%). The FosU_{PCA}-cJun complex (A and C, purple) shows increased helicity (fH = 39.8%), but a similar helical structure (222 nm/208 nm = 0.78), compared to the FosU_{PCA}-cFos (B and D, purple) complex (222 nm/208 nm = 0.81, fH = 27.6%) Thermal denaturation profiles with cJun (C) and cFos (D) were taken using 1 °C increments and tracking the 222 nm signal at 150 μ M. FosU_{PCA}-cJun shows an increase in the transition midpoint with a T_m of 52.0 °C compared to FosU_{PCA} in isolation and FosU_{PCA}-cFos, both with T_m unable to be fit. This suggests that FosU_{PCA} would preferentially bind to cJun over the potential off-target states. All experiments were performed in 10 mM potassium phosphate and 100 mM potassium fluoride (pH 7). Where possible (C), data were fitted to a two-state denaturation model.

FosU_{PCA} in isolation displayed a low level of helical stability (fH = ~17%) and a 222 nm/208 nm ratio of 0.46. In combination, this data describes a peptide lacking the prerequisite α -helical profile required for the formation of a homodimeric coiled coil. The inspection of the helical wheel (Figure 3a) shows the presence of six repulsive Glu-Glu interactions (60% of the possible total electrostatic interactions). Similarly, cFos (Figure 4b) exhibited low levels of helicity (fH = ~23%) and a 222 nm/208 nm ratio of 0.61. However, cJun displayed increased higher levels of helicity (fH = ~31%) and a 222 nm/208 nm ratio of 0.73 and is well-documented to be capable of forming a coiled coil. The secondary structure of the undesired FosU_{PCA}-cFos complex was analyzed using CD (Figure 4B) to establish if it formed from the component peptides. Although it presented with an fH of ~28%, the 222 nm/208 nm ratio of 0.81 also described a

structure tending toward an α -helix. Analysis of the target complex of FosU_{PCA}-cJun provided significant evidence for the formation of a coiled coil. In particular, it displayed increased α -helicity when compared to the homodimeric and off-target heterodimeric complexes (fH = ~40%) and four net electrostatic attractions (Figure 4b).

Structural analysis of complexes containing JunW_{CAND1} (Figure 5a/b) showed that the homodimeric complex

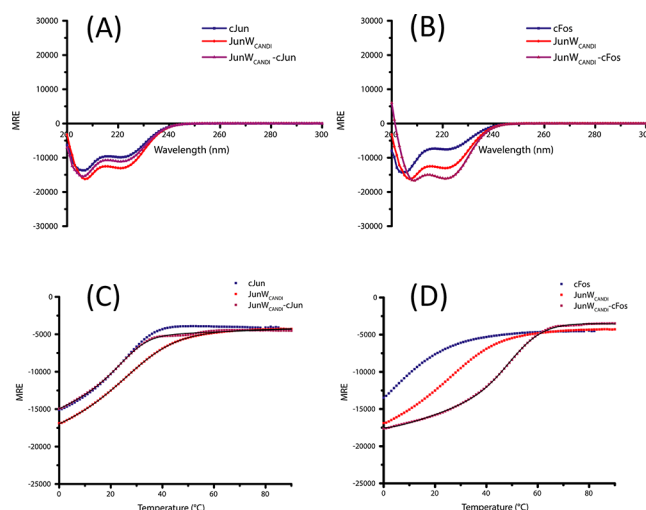


Figure 5. CD spectra and thermal denaturation data for JunW_{CAND1} with cJun and cFos. Shown are data for JunW_{CAND1} (red) with cJun (A and C, black) and cFos (B and D, blue). Spectra were measured at 20 °C at a total peptide concentration of 150 μ M and presented as mean residue ellipticity (MRE). The minima at 208 and 222 nm are indicative of a helical structure when coupled with fractional helicity (fH), with the 222 nm/208 nm ratio of JunW_{CAND1} showing stability in isolation (222 nm/208 nm = 0.81 and fH = 37.2%) greater than that of cJun (222 nm/208 nm = 0.73, fH = 30.5%). The JunW_{CAND1}-cJun complex (brown) shows helicity slightly greater than that of cJun (fH = 32.4%), but a similar helical structure (222 nm/208 nm = 0.74). In comparison, the JunW_{CAND1}-cFos complex (purple) showed an increase in helicity, with a 222 nm/208 nm ratio of 0.97 and a fH of 45.7%. Thermal denaturation profiles with cJun (C) and cFos (D) were taken using 1 °C increments and tracking the 222 nm signal at 150 μ M. JunW_{CAND1} shows a stable complex with a T_m of 32.0 °C compared to JunW_{CAND1}-cJun, with a slightly lower T_m of 27 °C. In complex with cFos, there is an increase in the transition midpoint with a T_m of 51 °C. This suggests that, although JunW_{CAND1} has stability in isolation and with cJun, it would preferentially bind to cFos over the potential off-target states. All experiments were performed in 10 mM potassium phosphate and 100 mM potassium fluoride (pH 7). Where possible (C), data were fitted to the two-state model.

presented with moderate levels of α -helicity (fH = ~37%) as well as a 222 nm/208 nm ratio of 0.81. In complex with cJun, the α -helicity showed only a slight decrease, with the fH value dropping to ~32% and the 222 nm/208 nm ratio to 0.74. When in complex with the cFos target, there is a marked increase in the α -helicity measured (fH = ~46%) and the 222 nm/208 nm ratio (0.97).

In context, these values describe a system in which homodimeric and off-target interactions retain relatively high levels of helical stability required to form the coiled-coil (with α -helical levels exceeded only by that of the JunW_{CAND1}-cFos interaction). FosU_{PCA} incubated with JunW_{CAND1} exhibited increased helicity compared to that of the component peptides (fH = ~52%) and a 222 nm/208 nm ratio of 0.89 (Figure 6b).

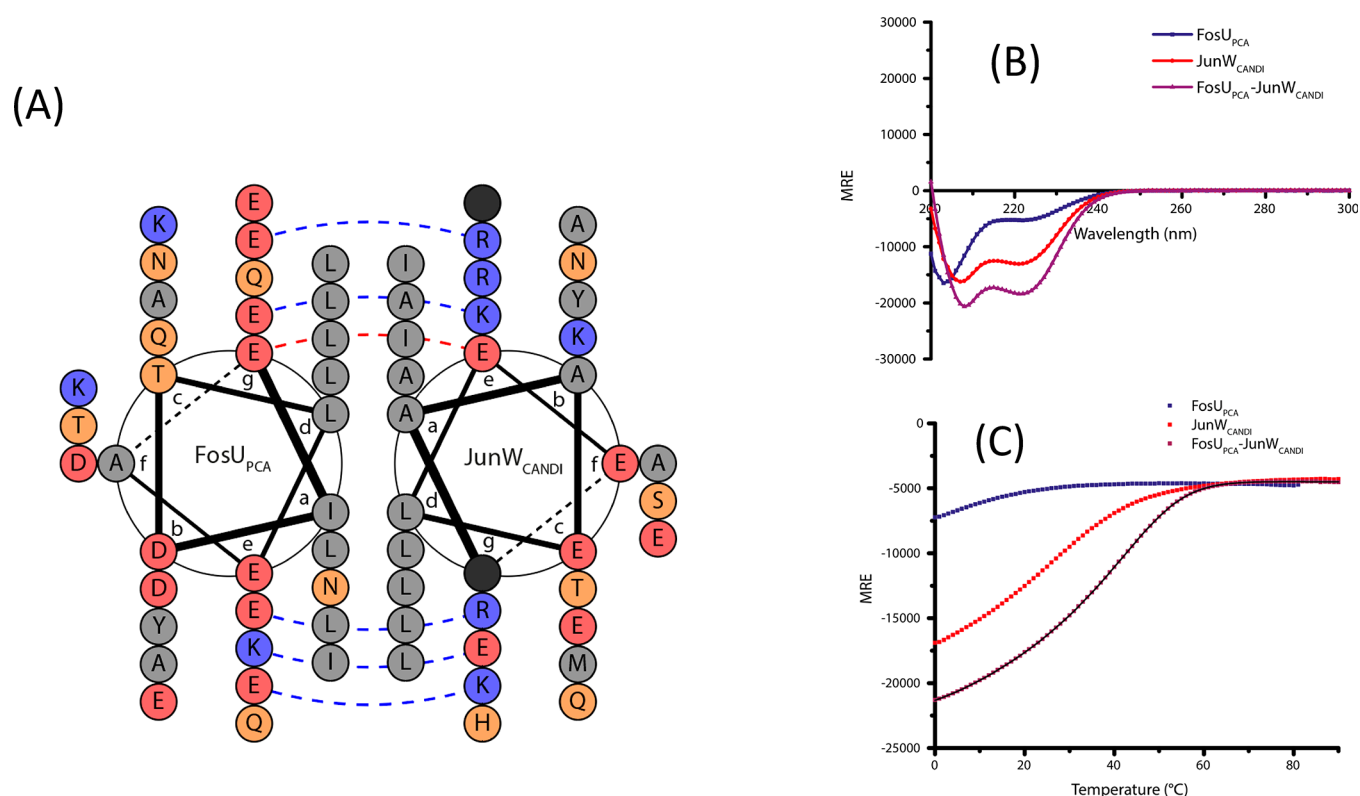


Figure 6. Helical wheel and CD spectra and thermal denaturation data for FosU_{PCA} and JunW_{CANDI}. The helical wheel (A) shows the amino acid arrangement for FosU_{PCA} and JunW_{CANDI} peptides. Spectra (B) were measured at 20 °C at a total peptide concentration of 150 μ M and presented as mean residue ellipticity (MRE). The minima at 208 and 222 nm are indicative of various levels of helical structure, with the 222 nm/208 nm ratio of the FosU_{PCA}–JunW_{CANDI} (purple) showing increased structure (222 nm/208 nm = 0.89) and helicity (α H = 51.9%). The lactamized form shows increased structure (222 nm/208 nm = 0.74) compared to the homomeric state (222 nm/208 nm = 0.69). Thermal denaturation profiles of FosU_{PCA}–JunW_{CANDI} (C), as well as the component peptides, were taken using 1 °C increments and tracking the 222 nm signal at 150 μ M. The heterodimer shows an increase in the transition midpoint, demonstrating a T_m of 40 °C compared to the only component able to have a T_m fitted (JunW_{CANDI} T_m = 32 °C). This suggests that the addition of JunW_{CANDI} promotes stability and helicity to FosU_{PCA} and that this complex is preferred over both homomeric states. This is addressed in the helical wheel, which shows a series of intermolecular attractive electrostatic interactions and repulsive intramolecular interactions. All experiments were performed in 10 mM potassium phosphate and 100 mM potassium fluoride (pH 7). Where possible (D), data were fitted to the two-state model.

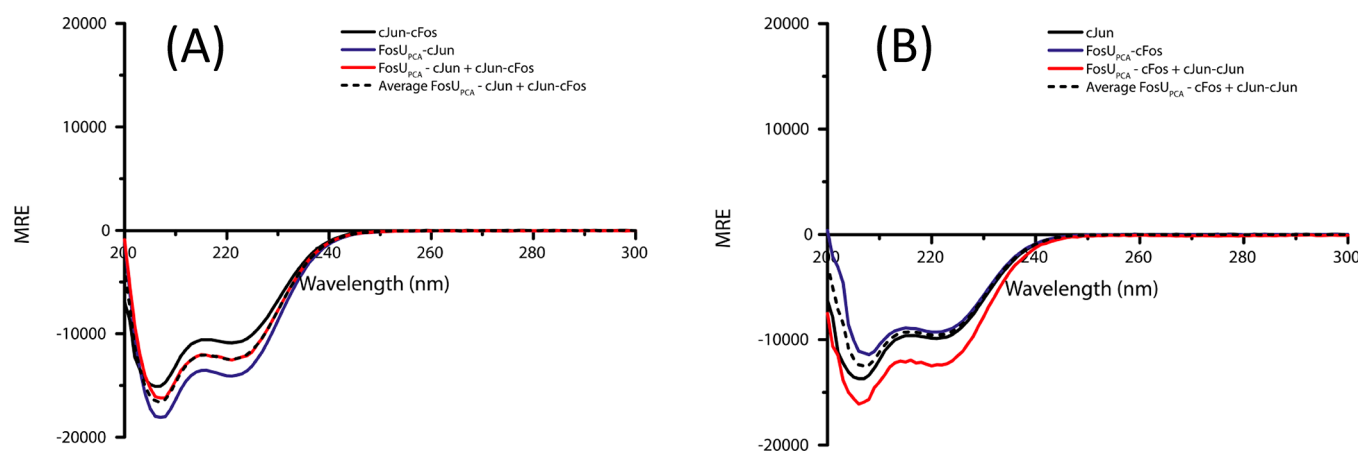


Figure 7. Dimer exchange with FosU_{PCA}, cJun, and cFos. (A) Equimolar mixtures of cJun–cFos and FosU_{PCA}–cJun were mixed, and the observed signal resembled the average of the two constituent spectra, indicating no change has occurred. (B) Equimolar mixtures of cJun–cJun and FosU_{PCA}–cFos were mixed, and the observed spectra exceeded the average of the two constituent spectra, which indicated that dimer exchange occurred to promote the system found in panel A. All experiments were performed at 150 μ M at 20 °C in 10 mM potassium phosphate and 100 mM potassium fluoride (pH 7).

Dimer Exchange. Dimer exchange experiments were performed for both the JunW_{CANDI} and FosU_{PCA} systems (Figures 7 and 8) in the presence of cJun and cFos cognate

binding partners. Upon mixing the four component peptides, the spectra generated demonstrate the global average (hashed line), should no exchange of the component peptides take

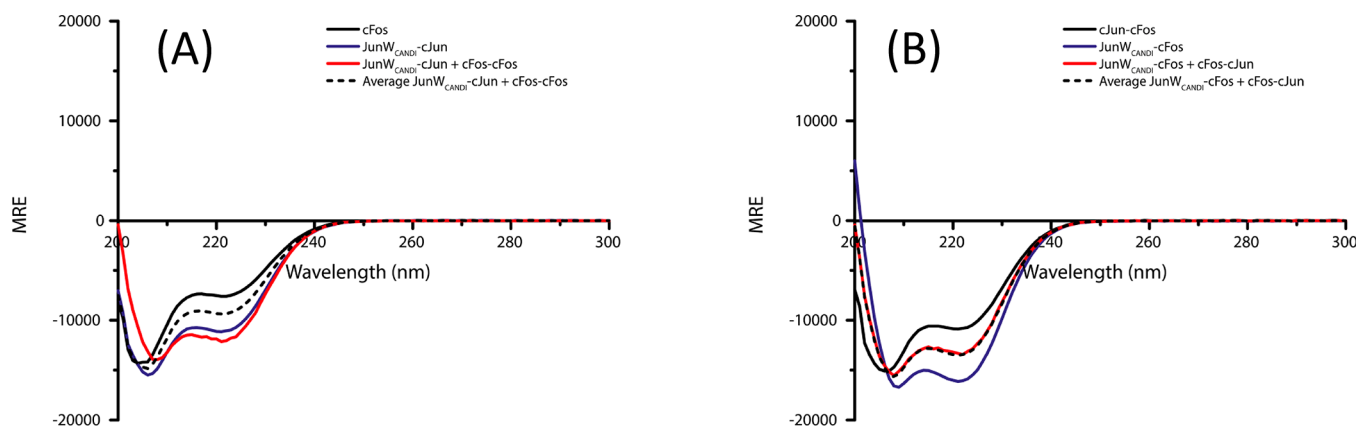


Figure 8. Dimer exchange with JunW_{CANDI}, cJun, and cFos. (A) Equimolar mixtures of cFos–cFos and JunW_{CANDI}–cJun were mixed, and the observed spectra exceeded that of the average, indicating change. (B) Equimolar mixtures of cJun–cFos and FosU_{PCA}–cFos were mixed, and the observed spectra resembled the average of the two constituent spectra, which indicated that no dimer exchange occurred. All experiments were performed at 150 μ M at 20 $^{\circ}$ C in 10 mM potassium phosphate and 100 mM potassium fluoride (pH 7).

place. No exchange was deemed to have occurred in systems containing FosU_{PCA}–cJun upon mixture with cJun–cFos (Figure 7a) as well as JunW_{CANDI}–cFos upon mixture with cFos–cFos (Figure 8b). In contrast, spectra exhibiting signals exceeding the average indicated that the expected dimer exchange had indeed occurred. These changes in binding partners were observed in systems containing off-target peptides in complex with either FosU_{PCA} (i.e., FosU_{PCA}–cFos combined with cJun–cJun; Figure 7b) or JunW_{CANDI} (i.e., JunW_{CANDI}–cJun combined with cFos–cFos; Figure 8a). These data provide evidence for the formation of hetero-specific coiled coils with the cognate AP-1 component, in the contact of alternative cJun or cFos options available. The data also further validate the results of the *in cellulo* screening step through PCA.

Thermal Denaturation Profiles. The increased global secondary structure content for both antagonist–target complexes required further stability validation through thermal denaturation experiments (Figure 4c/d and Figure 5c/d). Taken in 1 $^{\circ}$ C increments, the thermal melt data was generally in agreement with the spectra. In isolation, FosU_{PCA} did not form a stable coiled coil, with only the characteristic upper baseline observed (Figure 4c, red). Similarly, FosU_{PCA}–cFos showed a similar upper baseline only (Figure 4d, purple). For both of these complexes, the combination of spectra and thermal denaturation data demonstrates weakly populated helices without the ability to associate. This scenario is beneficial for this antagonist system, as it removes two off-target states, in addition to the observation that FosU_{PCA} incubated with cJun displayed a two-state sigmoidal thermal denaturation profile with a substantial right-hand shift ($T_m = 52$ $^{\circ}$ C). As shown from the spectra, JunW_{CANDI} in isolation (Figure 5c, red) formed a self-associating coiled coil with a T_m of 32 $^{\circ}$ C (similar to described previously¹¹). When incubated with cJun (Figure 5c, purple), only limited interaction was found to occur ($T_m = 27$ $^{\circ}$ C). Compared to thermal data with cJun, this represented an increase of 4 $^{\circ}$ C from that previously reported (JunW_{CANDI}–cJun₃₇ $T_m = 23$ $^{\circ}$ C). Incubated with cFos (Figure 5d, blue), as expected, a right-hand shift was observed ($T_m = 51$ $^{\circ}$ C), representing an increase of 7 $^{\circ}$ C from previous work with truncated cJun₃₇ ($T_m = 44$ $^{\circ}$ C). As observed in Figure 3e/f, these differences in T_m result from the addition of two pairs of g – e'^{+1} interactions from the mismatch in length.

Only in the off-target complex with cJun did this add an attractive electrostatic interaction (Glu–Arg at e^1 – g'^0), with the presence of other residues contributing solely to increasing helicity in the complex. JunW_{CANDI} in complex with FosU_{PCA} showed a right-hand shift compared to the component peptides (Figure 6c, purple). This increase in T_m to 40 $^{\circ}$ C, coupled with the spectra data, is evidence for a stable coiled coil. In context of the helical wheel (Figure 6a), this can be explained in part due to the fact that a net 4 of the 8 complete electrostatic interactions are attractive (Glu–Lys or Glu–Arg). In the context of the other T_m values, this stability does not affect the cocompatibility as it is lower than the desired target complex T_m values.

A comparison of the predicted and experimental T_m values (Figure 2 and Figure 9) shows notable differences, an observation that has previously been discussed in the context of bCIPA, isCAN, and this particular library.^{17,19}

Primarily, the off-target FosU_{PCA}–FosU_{PCA} and FosU_{PCA}–cFos interactions were predicted to be relatively stable (60 and 26 $^{\circ}$ C, respectively), whereas this was not observed experimentally. This highlights the role of the electrostatic interactions, as an analysis of the sequence (Figure 3a/c) shows that the homodimer has 60% of these as repulsive Glu–Glu interactions. FosU_{PCA}–cFos exhibits a similar profile, with 40% of these interactions as Glu–Glu. In addition, the presence of Leu at g^0 and e^4 as residues incapable of forming electrostatic interactions adds to this disruption, along with the presence of two Thr and two Lys residues at the core. Although less extreme, observation of the desired complex with cJun showed that there was also a discrepancy between experimental and predicted T_m values, with bCIPA predicting a T_m value of 86 $^{\circ}$ C and the experimental data displaying a T_m of 52 $^{\circ}$ C. Although this was a decrease of 34 $^{\circ}$ C, the system itself was not heavily altered. This is due to the fact that the ΔT_m decreased by 4 $^{\circ}$ C from a predicted 26 $^{\circ}$ C to an experimentally derived 22 $^{\circ}$ C. In comparison to previous work, this ΔT_m value is second only to that of FosU_{isCAN} by 8 $^{\circ}$ C. However, of the two, the FosU_{PCA} predicted ΔT_m value sits closer to that of the experimental value (FosU_{isCAN} $\Delta T_m = -20$ $^{\circ}$ C). In both cases, the predicted ΔT_m was driven by the difference in T_m between the desired complex with cJun and the homodimeric library member interaction. However, with the inability of either homodimer to form a stable coiled coil, this ΔT_m relied

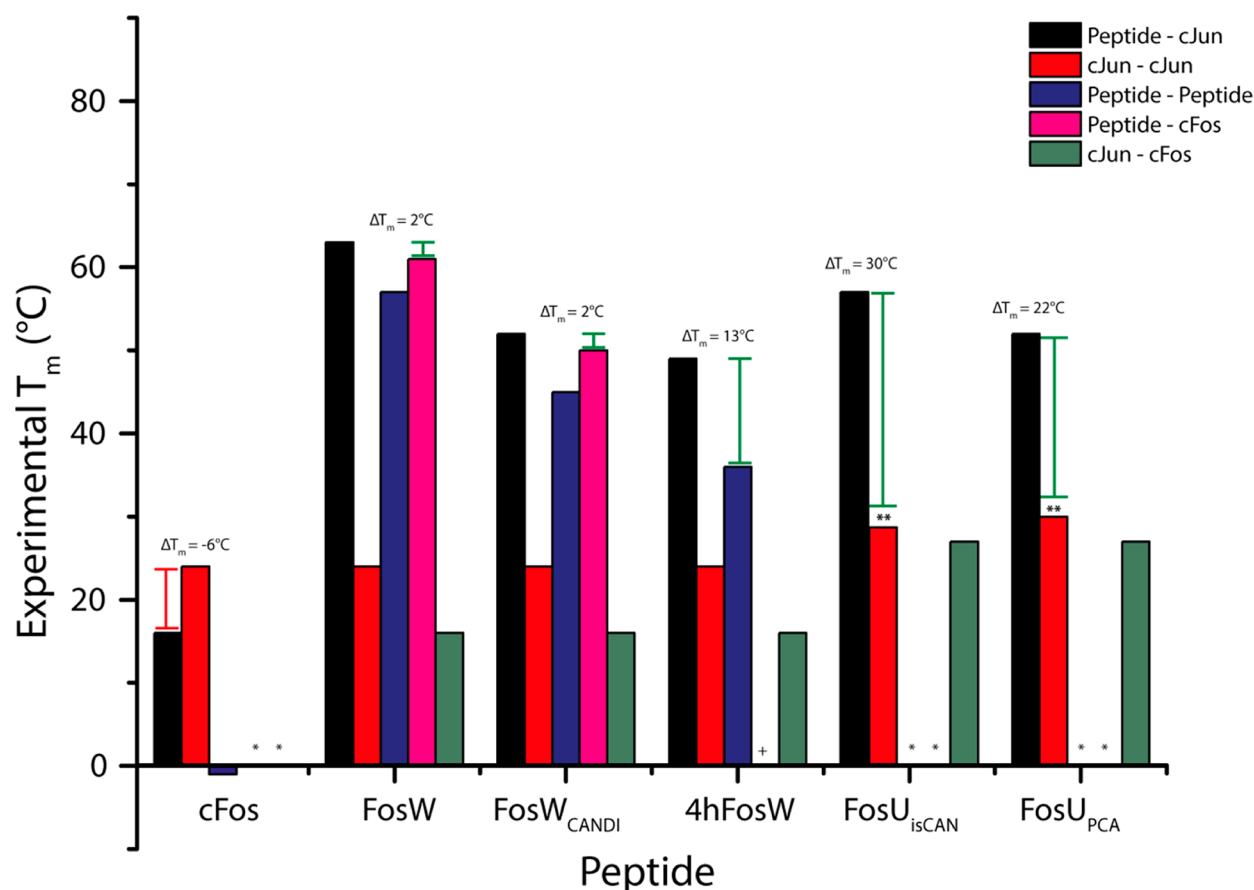


Figure 9. A comparison of measured FosU_{PCA} T_m data with previously designed peptides. The T_m of FosU_{PCA}-cJun is 52 °C with a ΔT_m of 22 °C. *FosU_{isCAN} and FosU_{PCA} did not form a stable homodimer or cFos heterodimer able to be fitted to get T_m . ⁺4hFosW-cFos lacks melt data. **cJun is an extended 39-mer sequence containing one extra g and one extra e position residue at the N-terminus and C-terminus, respectively, to add two extra interhelical electrostatic interactions with FosU_{isCAN} and FosU_{PCA}.

on the difference between the desired complex and the T_m of homodimeric cJun. The thermal stability of FosU_{PCA}-cJun is comparable to previous peptides for which extensive biophysical data is available.^{11,13,22,28} In particular, the PCA derived FosW-cJun exhibited a T_m value of 63 °C and a K_d value of 39 nM, whereas 4hFosW-cJun exhibited a T_m of 49 °C and a K_d value of 480 nM. Since the latter exhibited thermal stability within 3 °C of both FosU_{PCA}-cJun and JunW_{CANDI}-cFos, it can be estimated that the interaction K_d for both of these complexes is also within the nanomolar range.

Although the FosU_{PCA}-JunW_{CANDI} complex shows a high level of stability (T_m = 40 °C), it is important to note that JunW_{CANDI} was incorporated as a prototypical peptide sequence and neither peptide was explicitly screened (*in silico* or *in cellulo*) with the other as an off-target. The secondary structure and thermal data represent the formation of a stable complex, although this does not affect the cocompatibility of the two antagonists as an 11–12 °C difference exists between it and the T_m values of both FosU_{PCA}-cJun and JunW_{CANDI}-cFos. This can be credited to the fact that the increased electrostatic interactions in FosU_{PCA} and the stringency of the multipart *in silico* screening of Fos-based peptides resulted in a library that exhibited higher levels of specificity to cJun itself, rather than all members of the Jun family.

FosU_{PCA} Sequence Core Analysis. As shown in Figure 3, residue selection at the *a* positions resulted in alternations between Ile and Leu on *a*¹, *a*², and *a*⁵, with *a*¹ selecting the

former and *a*² and *a*⁵ selecting the latter. Through all stages of PCA, the successful sequences (with the option of both Asn and Ile) were selected for the Asn at this position, highlighting the importance of the interfacial Asn-Asn interaction in forming dimeric coiled coils and preventing higher-order oligomeric states.^{26,27} Not conforming to the structure or size of the otherwise ubiquitous residues found at this interface, His is thought to be ill-suited to the environment of the hydrophobic core at *a*⁴. Although not present in Jun/Fos proteins, His is present in the bHLH-ZIP cMax, along with Met, at sequential *d* positions without disrupting the interaction required for transcriptional activity.²⁹ When entering PCA, this residue was replaced at P1 with Leu. One potential reason could be its position within the coiled coil, which unlike cMax, is not located in the proximity of the N-terminus. As it occupies space directly between an Asn-Asn interaction, as well as a position within the center of the helix, the presence of the imidazole ring side chain could destabilize the α -helix and prevent successful coiled-coil formation.

FosU_{PCA} Sequence Electrostatic Analysis. As the electrostatic positions introduce the most diversity in options, it is interesting first to note the locations within the sequence that were not semirandomized within the library. Importantly, at positions *g*¹, *e*², *g*³, and *e*⁴, Glu was selected by both isPCA/isCAN, indicating that this residue is sufficient at these positions to satisfy all conflicting design requirements.¹⁷ However, although this has been briefly discussed previously,¹⁷

it is important to interrogate this in the context of the *in cellulo* screening of a large library. At each position, Glu performs a combinatorial destabilization and stabilization role, depending on the complex the peptide adopts. As shown in Figure 3a–c, the selected residues at these positions may, in some cases, appear counterintuitive. As residues at these positions interact within the coiled coil, this results in four repulsive Glu–Glu interactions in the homodimeric complex. When in complex with cJun, the presence of Glu serves to stabilize the coiled coil. $g^1-e'^2$, $e^2-g'^1$, and $g^3-e'^4$ interact beneficially in the form of 2 Glu–Lys interactions and one Glu–Arg interaction, respectively ($e^4-g'^3$ forms a Glu–Thr interaction). Since isCAN additionally considers interactions with cFos, the same destabilizing effect observed with the homodimer complex is selected, with repulsive Glu–Glu interactions (and a nonoptimal Glu–Leu interaction on $g^3-e'^4$). Previous work discussed the role of intramolecular interactions in the formation of the coiled coil.²³ Briefly, the presence of “solid-charge blocks”, a consecutive run of positively or negatively charged residues at either e and/or g positions within the heptad repeat, can result in intramolecular repulsion, thereby strengthening the interactions between the helices at these positions. Conversely, alternating \pm charges at these positions results in attractive intramolecular interactions, which serve to lower the effect of attractive/repulsive electrostatic interactions between helices. FosU_{PCA} contains 60% Glu at electrostatic positions (i.e., three small charge blocks), which may serve to guide intermolecular repulsion (homomeric and in complex with off-targets) and optimize the beneficial interactions when in complex with cJun. Although it disrupts a pattern of Glu residues at this in the heptad, g^2 benefits from the selection of Gln due to the presence of Ala at e'^3 . Within the context of the FosU_{PCA}–cJun heterodimer, g'^0 and g'^1 on cJun are Arg and Lys, respectively, with intramolecular repulsion, therefore potentially strengthening the intermolecular Arg–Glu and Lys–Glu interactions with FosU_{PCA} at positions e^1 and e^2 .

Prediction vs Experimental. As discussed previously, predictions using the bCIPA algorithm can vary in accuracy using designed peptide sequences.^{17,19} Elongation of designed sequences with the aim of maximizing potential interactions highlights a potential limitation in current predictive capabilities. Electrostatic charge blocks are also a sequence-specific effect, which is currently not considered. bCIPA scores each residue independently of others within a helix but scores for the specific interaction between helices, with only the latter taking advantage of the coupling data from the many natural peptides used to train the model. Additionally, these sequences are shorter than the sequences evaluated in this work, providing another possible explanation for the discrepancy. However, the ability to impose artificial limits through an imposed Δ parameter mitigates against this, replacing the subtle differences in desired/undesired predictions with predictably larger variations. This parameter functions as a necessary safeguard for the additional layers of complexity employed. Previous work exploring this³⁰ has also highlighted limitations of relying solely on coupling data.

CONCLUSIONS

In Silico Led In Cellulo Design. Of the 10 possible interactions within the intended system (Figure 10), both antagonists successfully bound their respective targets, while avoiding their cFos/cJun off-target complexes. Although a stable coiled coil, the FosU_{PCA}–JunW_{CANDI} complex is not

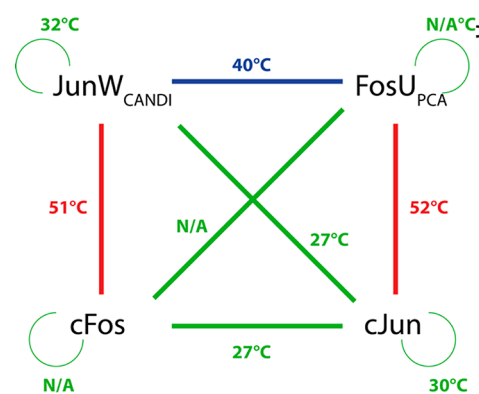


Figure 10. An overview of the FosU_{PCA} + JunW_{CANDI} system. Analysis of T_m data shows that the desired interactions (green) are preferentially formed over the off-targets (red) and the homodimers (black). Though the antagonist–antagonist dimer has a relatively high T_m (40 °C), this comparison highlights that the antagonist–target interactions are the preferred states.

predicted to disrupt the cocompatibility of the system due to the increased thermal stability of both of the peptides with their targets.

With work into FosU_{isCAN} and now FosU_{PCA}, respectively, exploring purely the data-driven and combinatorial *in silico/in cellulo* design, it has so far been shown that, aside from yielding increased stability of target homodimer complexes, the extension of the sequence length to add extra electrostatic interactions adds increased specificity but not increased stability in the heterodimeric peptide–target complex (Figure 7).¹⁷ Studies into the truncation of cJun antagonists have described similar issues.^{19,22} In these experiments, engineering stability has resulted in the antagonist peptide homodimers having near equivalent or increased thermal stability when compared to that of the desired heterodimeric complex with cJun. Due to the simplicity of coiled-coil sequences and the constraint imposed by the need to specifically target certain bZIPs without disrupting the function of others, this raises the question of whether there is a limit to the ability of certain engineered peptides to be able to fulfill both the stability and specificity design requirements. PCA-derived FosW exemplifies this phenomenon, with a difference of 6 °C between the T_m of its homodimeric complex and heterodimeric complex, in addition to the 2 °C between the latter and the T_m of a potential off-target target complex with cFos (Figure 7). As FosW–cJun, an interaction that did not consider cJun during design or selection, has a T_m that is only marginally higher than that of FosU_{PCA}–cJun, this highlights the advancement of *in silico* techniques in addressing the conflicting design requirements imposed by the simplicity of the coiled-coil structure. Though both FosW and FosU_{PCA} were derived via PCA, there was a significant difference in the size of the two libraries generated in order to do so (with the 49 152 member FosW library being 58% larger than that of the 31 104 member isCAN/isPCA). Despite this, in the latter case, the library derived a peptide that was more able to meet the criteria required by antagonist peptides, that is the maximization of stability balanced with the ability to specifically bind its target. In the context of engineering peptides, this represents an advancement in the process and validates the combined use of combinatorial *in silico* and *in cellulo* screening. To this end, future exploration into improving this framework would include an expansion of the library size as well as the

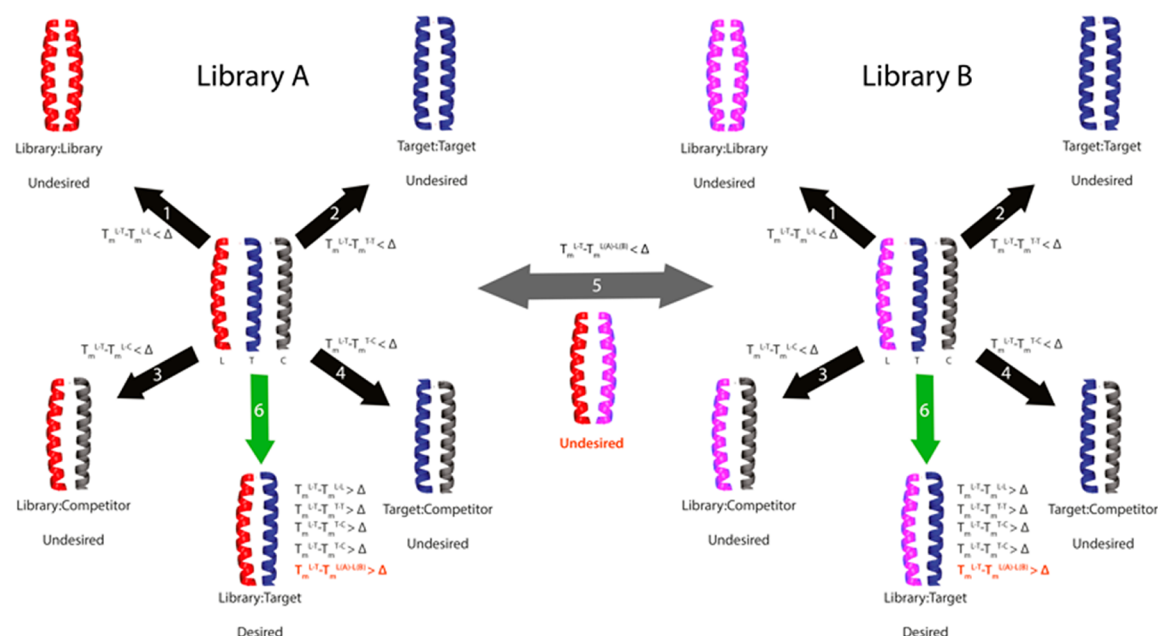


Figure 11. An overview of the proposed combined isCAN. This system incorporates the previously defined combination of predictions for off-target (1–4) and desired (6) complexes. It runs in parallel over two separate libraries (A/B) and allows for each member (L) of A to be incorporated as an additional competitor to the screening of library B and vice versa (5). This ensures that the libraries are capable of specificity relative to other designed peptides as driven by the Δ value (highlighted in red).

generation of an additional library containing cJun-based peptide sequences to target cFos.

As shown in Figure 11, the addition of that extra screening step (5) into this framework would allow not only for the isPCA/isCAN methods described here but also an additional isCAN constraint in which the cFos-based and cJun-based libraries are able to consider one another during the *in silico* screening stage. Similarly, the *in cellulo* screening would expand to incorporate a specific CANDI element, with both isPCA/isCAN derived peptides utilized as a competitor peptide during this process. This would allow for the library design and subsequent directed evolution of a potentially synergistic system, whereby two AP-1 inhibitors designed to target different components within an oncogenic heterodimer could function specifically with minimal cross-interactions.

In summary, this work provides a step forward in the use of a more stringent *in cellulo* screening, preceded by an *in silico* screening, to derive peptide antagonists for coiled-coil proteins in general, using AP-1 as an exemplar. The ability to utilize a combination of successful *in silico* and *in cellulo* screening methods has been validated by sequential combined screenings. Within the larger context, this represents progress in the ability to derive specific peptides capable of targeting key components within complex bZIP systems, while increasing the experimental data required to progress the data-driven design.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.biochem.9b00631>.

Overview of PCA passages, Fos library, and mass spectrometry data (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: j.mason@bath.ac.uk.

ORCID

Jody M. Mason: 0000-0002-4118-1958

Present Address

‡A.S.M.: AGC Biologics, Vandtaarnsvej 83B, 2860 Soeborg, Copenhagen, Denmark.

Author Contributions

A.S.M. and A.L. conducted the experiments and synthesized, purified, and characterized peptides and cJun. J.M.M. directed the research. A.L. and J.M.M. participated in experimental design, analysis of the data, and writing the paper.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

J.M.M. and A.L. thank the University of Bath for a Studentship. J.M.M. is also grateful to Cancer Research UK (A11738 and A26941) and to the BBSRC (BB/R017956/1) and EPSRC (EP/M001873/2).

■ REFERENCES

- (1) Passegue, E. (2000) JunB suppresses cell proliferation by transcriptional activation of p16 INK4a expression. *EMBO J.* 19, 2969–2979.
- (2) Milde-Langosch, K. (2005) The Fos family of transcription factors and their role in tumourigenesis. *Eur. J. Cancer* 41, 2449–2461.
- (3) Gupta, S., Kumar, P., Kaur, H., Sharma, N., Saluja, D., Bharti, A. C., and Das, B. C. (2015) Selective participation of c-Jun with Fra-2/c-Fos promotes aggressive tumor phenotypes and poor prognosis in tongue cancer. *Sci. Rep.* 5 (16811), 1–15.
- (4) Darnell, J. E. (2002) Transcription factors as targets for cancer therapy. *Nat. Rev. Cancer* 2, 740–749.
- (5) Eferl, R., and Wagner, E. F. (2003) AP-1: A double-edged sword in tumorigenesis. *Nat. Rev. Cancer* 3, 859–868.

- (6) Eckert, R. L., Adhikary, G., Young, C. A., Jans, R., Crish, J. F., Xu, W., and Rorke, E. A. (2013) AP1 Transcription Factors in Epidermal Differentiation and Skin Cancer. *J. Skin Cancer* 2013, 1–9.
- (7) Kappelman, M., Bosserhoff, A., and Kuphal, S. (2014) AP-1/c-Jun transcription factors: Regulation and function in malignant melanoma. *Eur. J. Cell Biol.* 93, 76–81.
- (8) Crick, F. H. C. (1953) The packing of α -helices: simple coiled-coils. *Acta Crystallogr.* 6, 689–697.
- (9) Fong, J., Keating, A., and Singh, M. (2004) Predicting specificity in bZIP coiled-coil protein interactions. *Genome Biol.* 5, R11.
- (10) Grigoryan, G., and Keating, A. E. (2006) Structure-based prediction of bZIP partnering specificity. *J. Mol. Biol.* 355, 1125–1142.
- (11) Mason, J. M., Müller, K. M., and Arndt, K. M. (2007) Positive aspects of negative design: Simultaneous selection of specificity and interaction stability. *Biochemistry* 46, 4804–4814.
- (12) Grigoryan, G., Reinke, A. W., and Keating, A. E. (2009) Design of protein-interaction specificity affords selective bZIP-binding peptides. *Nature* 458, 859–864.
- (13) Mason, J. M., Schmitz, M. A., Muller, K. M., and Arndt, K. M. (2006) Semirational design of Jun-Fos coiled coils with increased affinity: Universal implications for leucine zipper prediction and design. *Proc. Natl. Acad. Sci. U. S. A.* 103, 8989–8994.
- (14) Ye, N., Ding, Y., Wild, C., Shen, Q., and Zhou, J. (2014) Small molecule inhibitors targeting activator protein 1 (AP-1). *J. Med. Chem.* 57, 6930–6948.
- (15) Yin, Z., Machius, M., Nestler, E. J., and Rudenko, G. (2017) Activator protein-1: Redox switch controlling structure and DNA-binding. *Nucleic Acids Res.* 45, 11425–11436.
- (16) Tewari, D., Nabavi, S. F., Nabavi, S. M., Sureda, A., Farooqi, A. A., Atanasov, A. G., Vacca, R. A., Sethi, G., and Bishayee, A. (2018) Targeting activator protein 1 signaling pathway by bioactive natural agents: Possible therapeutic strategy for cancer prevention and intervention. *Pharmacol. Res.* 128, 366–375.
- (17) Lathbridge, A., and Mason, J. M. (2018) Computational Competitive and Negative Design to Derive a Specific cJun Antagonist. *Biochemistry* 57, 6108–6118.
- (18) Hagemann, U. B., Mason, J. M., Müller, K. M., and Arndt, K. M. (2008) Selectional and Mutational Scope of Peptides Sequestering the Jun-Fos Coiled-Coil Domain. *J. Mol. Biol.* 381, 73–88.
- (19) Lathbridge, A., and Mason, J. M. (2019) Combining Constrained Heptapeptide Cassettes with Computational Design To Create Coiled-Coil Targeting Helical Peptides. *ACS Chem. Biol.* 14, 1293–04.
- (20) Acharya, A., Rishi, V., and Vinson, C. (2006) Stability of 100 homo and heterotypic coiled-coil a-a' pairs for ten amino acids (A, L, I, V, N, K, S, T, E, and R). *Biochemistry* 45, 11324–11332.
- (21) Krylov, D., Barchi, J., and Vinson, C. (1998) Inter-helical interactions in the leucine zipper coiled coil dimer: pH and salt dependence of coupling energy between charged amino acids. *J. Mol. Biol.* 279, 959–972.
- (22) Crooks, R. O., Rao, T., and Mason, J. M. (2011) Truncation, randomization, and selection: Generation of a reduced length c-jun antagonist that retains high interaction stability. *J. Biol. Chem.* 286, 29470–29479.
- (23) Crooks, R. O., Lathbridge, A., Panek, A. S., and Mason, J. M. (2017) Computational Prediction and Design for Creating Iteratively Larger Heterospecific Coiled Coil Sets. *Biochemistry* 56, 1573–1584.
- (24) Fields, G. B., and Noble, R. L. (1990) Solid phase peptide synthesis utilizing 9-fluorenylmethoxycarbonyl amino acids. *Int. J. Pept. Protein Res.* 35, 161–214.
- (25) Mason, J. M., Hagemann, U. B., and Arndt, K. M. (2007) Improved stability of the Jun-Fos activator protein-1 coiled coil motif: A stopped-flow circular dichroism kinetic analysis. *J. Biol. Chem.* 282, 23015–23024.
- (26) Fletcher, J. M., Bartlett, G. J., Boyle, A. L., Danon, J. J., Rush, L. E., Lupas, A. N., and Woolfson, D. N. (2017) N@a and N@d: Oligomer and Partner Specification by Asparagine in Coiled-Coil Interfaces. *ACS Chem. Biol.* 12, 528–38.
- (27) Thomas, F., Niitsu, A., Oregioni, A., Bartlett, G. J., and Woolfson, D. N. (2017) Conformational Dynamics of Asparagine at Coiled-Coil Interfaces. *Biochemistry* 56, 6544–6554.
- (28) Baxter, D., Ullman, C. G., Frigotto, L., and Mason, J. M. (2017) Exploiting Overlapping Advantages of in Vitro and in Cellulo Selection Systems to Isolate a Novel High-Affinity cJun Antagonist. *ACS Chem. Biol.* 12, 2579–2588.
- (29) Battey, J., Moulding, C., Taub, R., Murphy, W., Stewart, T., Potter, H., Lenoir, G., and Leder, P. (1983) The human c-myc oncogene: Structural consequences of translocation into the igh locus in Burkitt lymphoma. *Cell* 34, 779–787.
- (30) Potapov, V., Kaplan, J. B., and Keating, A. E. (2015) Data-Driven Prediction and Design of bZIP Coiled-Coil Interactions. *PLoS Comput. Biol.* 11, e1004046.

5.5 Supporting Information

	g0	a1	b1	c1	d1	e1	f1
P0	Q	L	D	T	L	Q	A
P1	E	I	D	T	L	E	A
P2	E	I	D	T	L	E	A
Options	EKQ	IL				EKQ	

	g1	a2	b2	c2	d2	e2	f2
P0	E	I	D	Q	L	E	D
P1	E	I	D	Q	L	E	D
P2	E	L	D	Q	L	E	D
Options		IL					

	g2	a3	b3	c3	d3	e3	f3
P0	K	N	Y	A	L	K	T
P1	K	N	Y	A	L	Q	T
P2	Q	N	Y	A	L	K	T
Options	EKQ	IN				EKQ	

	g3	a4	b4	c4	d4	e4	f4
P0	E	H	A	N	L	E	K
P1	E	L	A	N	L	E	K
P2	E	L	A	N	L	E	K
Options		ILN(H)					

	g4	a5	b5	c5	d5	e5
P0	Q	I	E	K	L	Q
P1	Q	L	E	K	L	Q
P2	E	I	E	K	L	Q
Options	EQ	IL				EKQ

Table 1. Overview of PCA passages. The options (red) were randomised at 11 positions and the residue selections for each stage shown in blue. At both a^3 and e^5 , there were no changes in residue from the first passage (Asn and Gln, respectively).

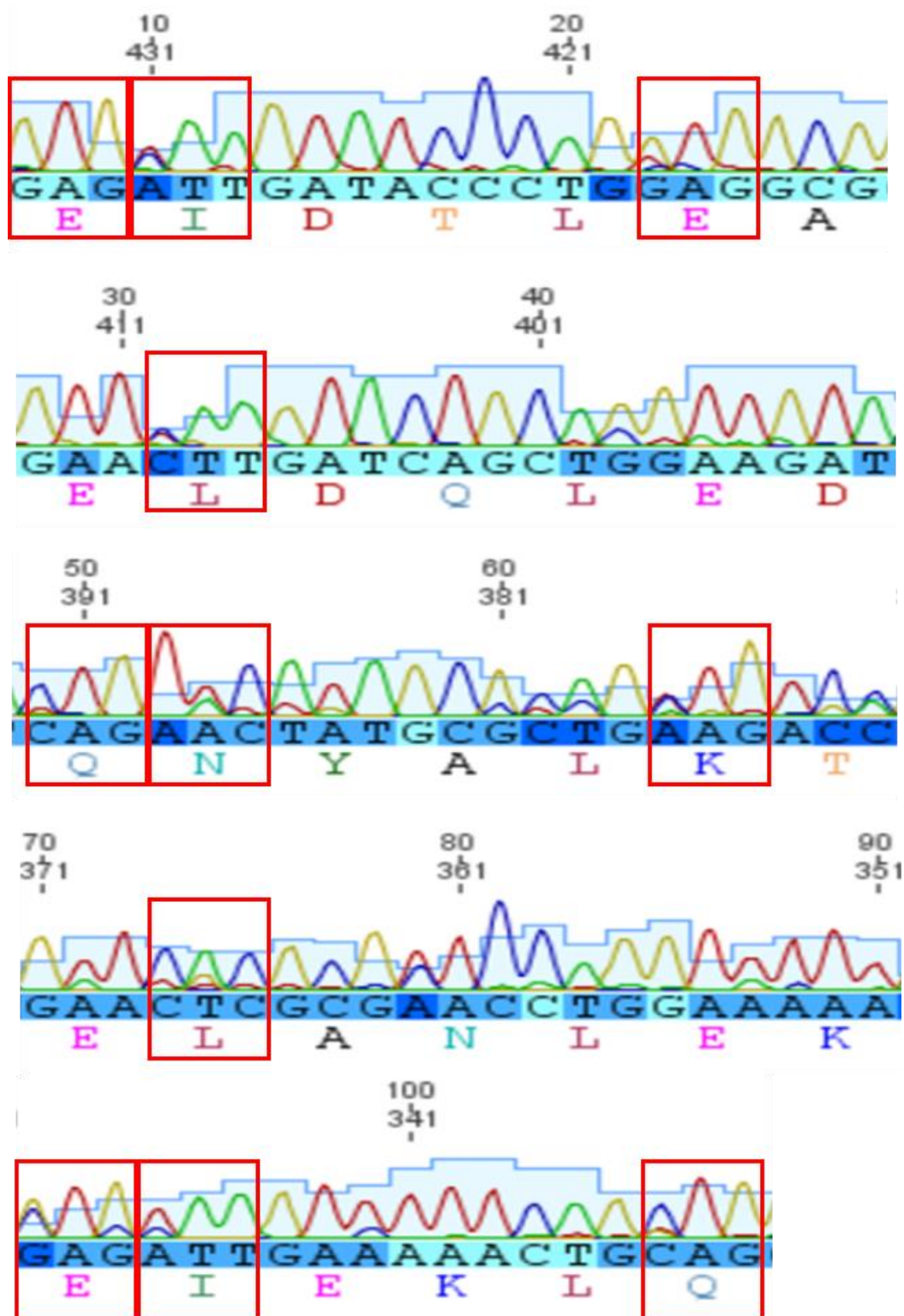
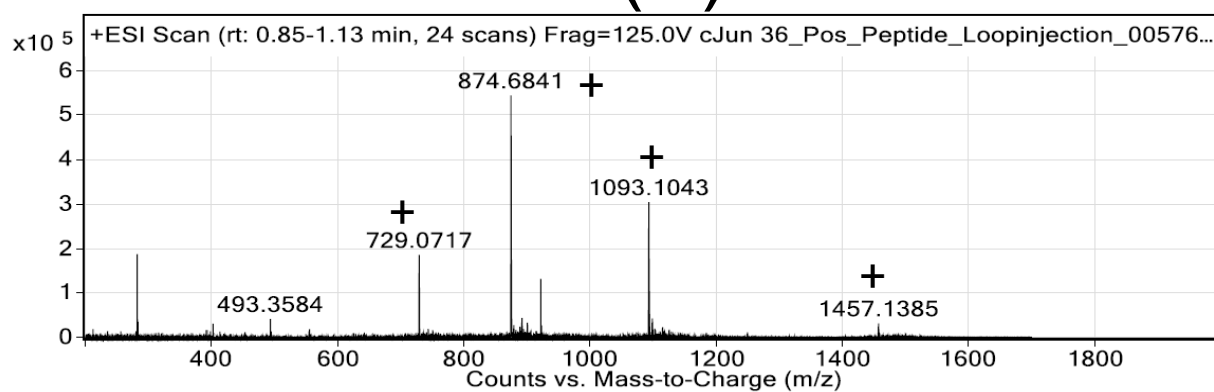


Figure S1. Fos Library Chromatogram. Sequence data from the second (and final) passage (P2) showed the selection results from 11 randomised positions in the sequence (highlighted in red). Glu residues were selected at g^0 , e^1 , and g^4 whereas Lys was selected at e^3 – with Gln residues on g^2 and e^5 . Ile was chosen at a^1 and a^5 , Leu at a^2 and a^4 , and Asn at a^3 .

(A)



(B)

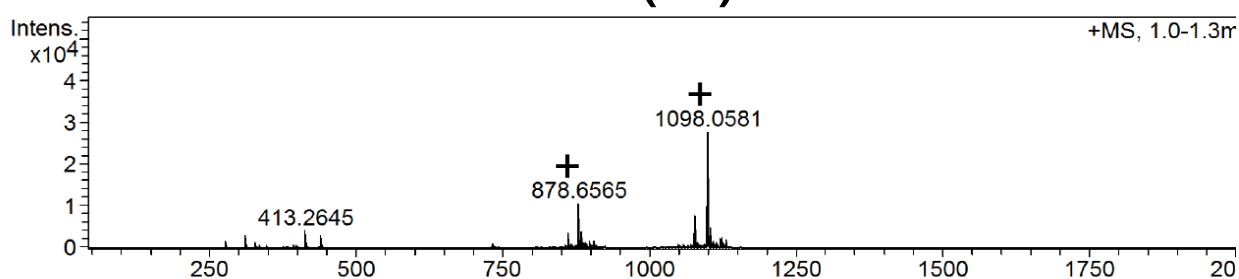
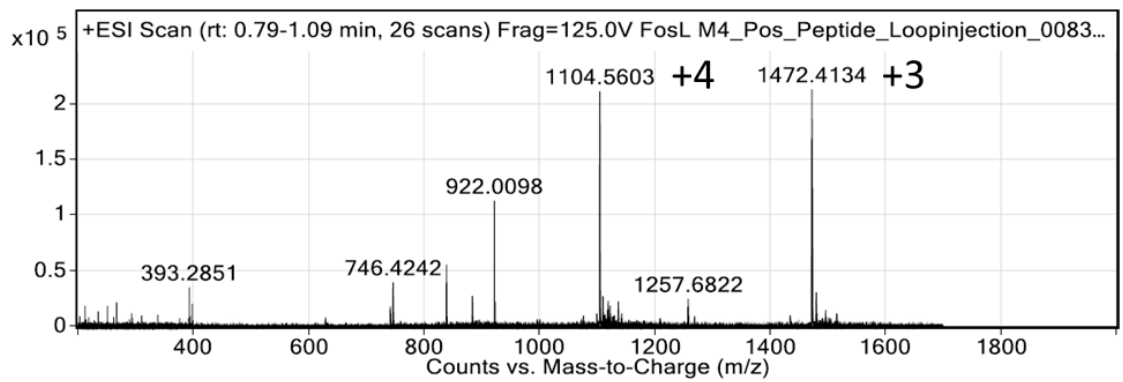


Figure S2. Electrospray Mass Spectrometry data from purified cJun and cFos demonstrating correct +3 - +5 m/z of cJun (A) and cFos (B)

(A)



(B)

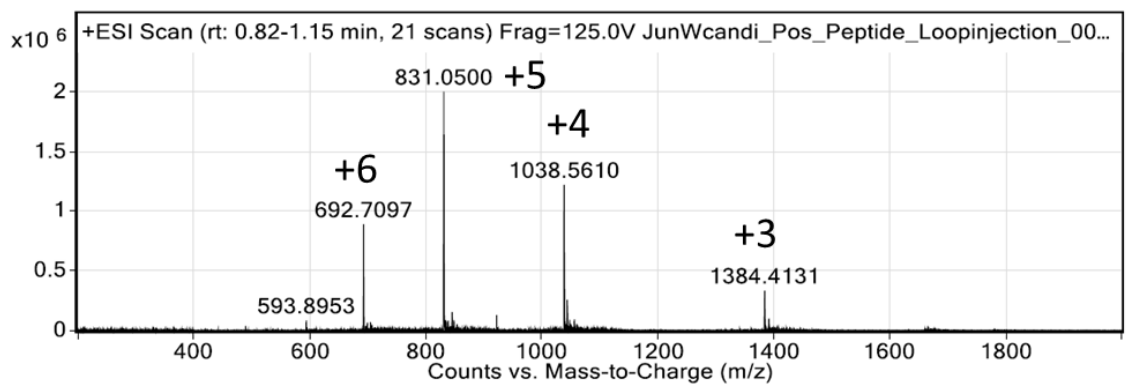


Figure S3. Electrospray Mass Spectrometry data from purified FosU_{PCA} and JunW_{CANDI} demonstrating correct +3 - +5 m/z of FosU_{PCA} (A) and JunW_{CANDI} (B)

5.6 Concluding Remarks on Chapter 5

As discussed in Chapter 4, this addition of PCA further demonstrated the capability of *in-silico* design of a cJun antagonist. A key element of this was the ability of the *in-silico* steps to add an extra layer of screening to the overall library development.

Similar to the work within Chapter 4 (Lathbridge and Mason 2018), the minimisation of the library size within this study did not result in a less successful antagonist in comparison to previous sequences taken from larger *in-cellulo* libraries (Mason et al. 2006; Mason et al. 2007).

Though undeniably powerful, the work so far within this thesis has highlighted a limitation in the technique. As we move further into exploring elongated peptides, the accuracy of the underlying bCIPA is reduced - explaining the discrepancy between the predicted T_m values and the experimental values. This is due to the fact that the algorithm itself was trained on sequences that were not as long nor as optimised as those behind explored currently.

The *in silico* screening works on the principle that the hydrophobic core and electrostatic predictor data is not affected by the local residues at other positions (while also assuming that the helical contributions of residues remain constant regardless of location within the helix). As has been shown, the selection of similarly charged residues to form a “solid charge block” has been a pattern within isCAN and the potential intra-helical effects on stability may contribute to the discrepancy seen within T_m data.

Although this is the case, the work covered in the last 2 chapters has shown that the underlying methodology is sound and has allowed for the selection of sequences that favour the desired dimer state and disfavour the multiple off-target possibilities.

In this way, it is likely that further training of the algorithm to include sequences developed since its inception would result in different weighting given to the predictors through altered coefficient values.

Similarly to elongation, truncation of peptides has not been adequately explored in the context of bCIPA-based screening. As it is thought that the loss of both length and inter-helical interactions would decrease the accuracy of the approach, further work would be possible to look at the effect of helix stabilisation – thereby removing one of the predictors from the bCIPA calculation.


CHAPTER 6 - COMBINING CONSTRAINED HEPTAPEPTIDE CASSETTES WITH COMPUTATIONAL DESIGN TO CREATE COILED-COIL TARGETING HELICAL PEPTIDES

Reproduced with permission from:

Lathbridge, A., and Mason, J. M. (2019) Combining Constrained Heptapeptide
Cassettes with Computational Design To Create Coiled-Coil Targeting Helical
Peptides. *ACS Chemical Biology* 14 (6), 1293-1304

Copyright 2019 American Chemical Society.

Appendix 6B: Statement of Authorship

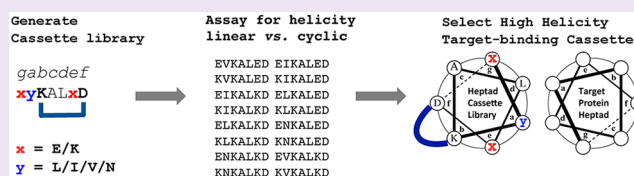
This declaration concerns the article entitled:			
Combining Constrained Heptapeptide Cassettes with Computational Design To Create Coiled-Coil Targeting Helical Peptides			
Publication status (tick one)			
Draft manuscript	<input type="checkbox"/>	Submitted	<input type="checkbox"/>
		In review	<input type="checkbox"/>
		Accepted	<input type="checkbox"/>
		Published	<input checked="" type="checkbox"/>
Publication details (reference)	Lathbridge, A., and Mason, J. M. (2019) Combining Constrained Heptapeptide Cassettes with Computational Design To Create Coiled-Coil Targeting Helical Peptides. <i>ACS Chemical Biology</i> 14 (6), 1293-1304		
Copyright status (tick the appropriate statement)			
I hold the copyright for this material		<input type="checkbox"/>	Copyright is retained by the publisher, but I have been given permission to replicate the material here
Candidate's contribution to the paper (provide details, and also indicate as a percentage)		<p>The candidate contributed to / considerably contributed to / <u>predominantly executed</u> the...</p> <p>Formulation of ideas:</p> <p>50% - Contributed to computational and <i>in vitro</i> ideas</p> <p>Design of methodology:</p> <p>80% - Developed software, cassette framework, and SPSS</p> <p>Experimental work:</p> <p>100% - Screening, characterisation, and data analysis</p> <p>Presentation of data in journal format:</p> <p>80% - Contributed to manuscript</p>	
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.		
Signed			Date 10/06/19

Combining Constrained Heptapeptide Cassettes with Computational Design To Create Coiled-Coil Targeting Helical Peptides

Alexander Lathbridge[†] and Jody M. Mason^{*,†,‡,§}[†]Dept of Biology and Biochemistry and [‡]Centre for Therapeutic Innovation, University of Bath, Claverton Down, Bath BA2 7AY, United Kingdom

S Supporting Information

ABSTRACT: A total of 32 heptapeptides have been synthesized and characterized to establish the effect of K → D ($i \rightarrow i + 4$) lactamization upon their ability to adopt a helical conformation. Because most parallel and dimeric coiled-coil sequences can be deconvoluted into *gabcdef* repeats, we have introduced fixed solvent exposed b → f (K → D) constraints into this design scaffold. Interfacial “a” hydrophobic (L/I/V/N) and “e/g” electrostatic (E/K) options ($4 \times 2 \times 2 = 16$ cassettes) were introduced as core drivers of coiled-coil stability and specificity. All present as random coils when linear but adopt a helical conformation upon lactamization. Helicity varied in magnitude from 34 to 68%, indicating different levels of constraint tolerance within the context of a sequence required to be helical for function. Using the oncogenic transcription factor cJun as an exemplar, we next utilized our bCIPA coiled-coil screening engine to select four cassettes of highest predicted affinity when paired with four *gabcdef* cassettes within the full-length cJun target counterpart ($16^4 = 65\,536$ combinations). This information was coupled with observed helicity for each constrained cassette to select for the best balance of predicted affinity when linear and experimentally validated helicity when constrained. As a control, the same approach was taken using cassettes of high predicted target affinity but with lower experimentally validated helicity. The approach provides a novel platform of modular heptapeptide cassettes experimentally validated and separated by helical content. Appropriate cassettes can be selected and conjugated to produce longer peptides, in which constraints impart appropriate helicity such that a wide range of targets can be engaged with high affinity and selectivity.



It is very difficult to establish if a given peptide sequence will tolerate a helix-inducing constraint, meaning that their introduction into peptide sequences toward the goal of imparting helicity, and ultimately increased target affinity, is largely a trial-and-error process.¹ However, it has long been known that peptides of less than 15 residues (1–4 helical turns) are unable to independently form thermodynamically stable α -helices in water.² Longer helices form due to a sequential development of intrastrand hydrogen bonds, propagating from a N-terminal folding nucleus toward the C-terminus.³ This process is interrupted within short synthetic sequences as water molecules compete with the C=O:H–N interactions of the peptide backbone, meaning that the formation of the helix becomes energetically less favorable, because the entropic cost associated with folding increases.⁴ Helical stabilization of shorter peptides is therefore an area of intense interest, particularly in protein–protein interactions where biological activity is mediated by helicity.⁵ Multiple methods to modify peptides to increase their stability have been explored, including hydrogen bond surrogates,^{6,7} triazole linkers,⁸ hydrocarbon staples,⁴ double-click linkers,^{9,10} lactam bridges,^{2,9–12} and other techniques based on macrocyclization chemistries.^{9,10}

Of all helix-inducing constraints currently available, lactamization is a particularly powerful approach to increasing the

stability of short helical peptides. It is an example of side-chain stapling, in which, for example, a peptide bond can be formed between Lys and Asp residues. By coupling amine and carboxylate containing side-chains at residues spaced four residues apart ($i \rightarrow i + 4$ configuration), the constraint is able to function such as to complement that of the hydrogen bonds found in native α -helices.^{4,5,9,13} Although successful lactamization of small peptides has been observed,^{11,14–17} there is still a lack of comparative data on tolerance with respect to longer peptide sequences, and consequent designs that can be implemented toward practical usage. Our group have adopted a strategy of K → D lactamization, because this has been shown to be the most potent inducer of α -helicity in short peptide sequences relative to alternative approaches.²

Here we describe a technique that combines computational design^{18–22} with single-heptad lactamization to form stable and functional coiled-coil peptides. We are primarily focused on the comparison of $i \rightarrow i + 4$ (K → D) lactamization tolerance within individual heptad cassettes corresponding to coiled-coil sequences. This is toward a major goal of increasing helicity conferred upon parental sequences, in which heptad repeats

Received: April 4, 2019

Accepted: May 21, 2019

Published: May 22, 2019

are contained, thus avoiding a trial-and-error search for constraints that will be tolerated or that, more importantly, will impart helicity. We are additionally interested in combining the most promising lactamized heptads to form longer peptides, in which they are tolerated. Previous studies of helicity imposed by lactamization of pentapeptides have shown that this has the potential to be used as a generalized approach.¹² We have previously demonstrated the *bZIP Coiled coil Interaction Prediction Algorithm* (bCIPA) to be a valuable tool in the design of peptides that can antagonize coiled-coil interactions involved in the formation of basic leucine zippers (bZIPs).^{18,21,22} Here we describe the combination of computational and experimental techniques toward investigating the ability to generate helically constrained heptads. These are then applied toward the larger goal of designing longer constrained α -helices, in which the likelihood of achieving increased helicity and more importantly, improved binding, is increased. As a proof of concept, we explore the ability of modular heptapeptide cassettes that are capable of interfering with the native cJun coiled coil, relative to a counterpart that despite computational selection, contains heptapeptide modules that are found to constrain poorly. Specifically, we seek to address: (i) Does lactamization promote helicity in heptapeptide cassette sequences? (ii) Does the helicity imparted in any given cassette vary according to the sequence? (iii) Are cassettes able to be conjugated into longer sequences while retaining/improving helicity relative to linear counterparts? (iv) Using an exemplar target, can such sequences be used to promote increased target affinity relative to their linear counterparts?

RESULTS AND DISCUSSION

There are a range of methodologies for constraining short peptides into helical structures, but the most potent inducer of helicity for short sequences has been shown to be $K \rightarrow D$ ($i \rightarrow i + 4$) lactamization.^{15,23} Lactamization has the added benefit of introducing a very discrete change to the linear peptide sequence; the constraint consists simply of a condensation reaction between two naturally occurring side-chain sequences, resulting in the formation of a peptide bond. Despite this, for a given sequence, it can be very difficult to predict if a lactam bridge will be tolerated, meaning that the process of their introduction to impart increased helicity upon a peptide is trial and error. Here we describe the first use of an approach that seeks to synthesize and characterize modular cassettes with a wide range of sequence-specific properties that are desired for coiled-coil forming interactions, such that they can be sorted into (i) those of experimentally validated high helicity and (ii) those predicted to bind to an oncongenic coiled-coil target sequence. The study provides the first set of modular cassettes that can serve such a purpose, with a view to being able to select the most appropriate balance of helicity and target affinity, such that cassettes can be conjugated toward targeting the cJun coiled-coil target.

Heptad Library Design. We have created a peptide library consisting of seven residue sequences that correspond to one heptad repeat of a coiled-coil motif (*gabcdef; gaKALeD*). In this library, positions *g* and *e*, which are important in forming electrostatic contacts within a coiled-coil sequence,^{24,25} were semi-randomized to generate E/K options, with a view to generating potential attractive and repulsive options with the corresponding positions of the target. Similarly, the *a* position corresponding to the core region within a coiled-coil sequence

was semi-randomized to generate L/I/V/N options. The *c* and *d* positions were fixed as A and L, respectively, to impart helicity and further core hydrophobicity that is characteristic of the parallel dimeric coiled-coil motif.²⁶ Each peptide was next synthesized in both linear and lactamized form ($b \rightarrow f$; $K \rightarrow D$), to probe for constraint tolerance and helical induction (Figure 1). The options for each cassette therefore provided a

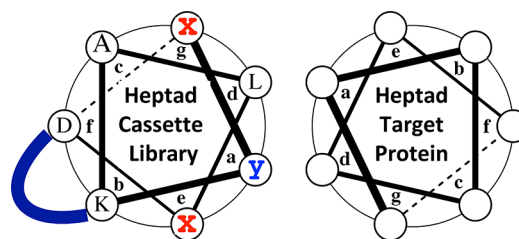


Figure 1. Design of heptad cassette sequences. Peptide options are semi-randomized around electrostatic positions *g* and *e* ($x = \text{Glu/Lys}$) and core position *a* ($y = \text{Leu/Ile/Val/Asn}$). This illustrates the diverse options available at the hydrophobic interface and the different charge profiles available for electrostatic interactions. The lactam bridge between b – f ($K \rightarrow D$) is also shown.

library of 16 constrained ($g/a/e$ options; $2 \times 4 \times 2 = 16$) and 16 linear heptapeptides with diverse electrostatic and hydrophobic characteristics. The options were provided to create a series of modular peptides, generating the required electrostatic and hydrophobic contributions to binding with any heptad counterpart within a defined target helix. In particular, by confining electrostatic interactions between peptide and target (i.e., $g_i - e_{i+1}$) to the same *gabcdef* repeat, interactions could be calculated completely independently during prediction, and therefore, each cassette could be considered truly modular. The final two solvent exposed positions (*b* and *f*; $i \rightarrow i + 4$ spacing) were chosen as the most appropriate positions to constrain within each modular cassette while avoiding potential interference with the binding interface.

Individual Cassette Helicity Measurements. Having synthesized all 16 peptides as both linear and constrained cassettes (see Figures S1–S4 for mass spectrometry data), we next sought to establish the extent to which each adopted a helical conformation upon introduction of the lactam constraint (Figure 2). As expected, all 16 sequences adopted a random coil confirmation in the absence of a lactam constraint, with one characteristic minimum at ~ 190 nm. In contrast, all 16 cassettes adopted an α -helical conformation upon introduction of the $K \rightarrow D$ constraint, with characteristic minima at 208 and 222 nm. However, the extent to which each of the 16 constrained cassettes adopted a helical conformation varied widely, with values ranging from 34 to 68% helicity (eq 1). The helicity of peptides in both linear and constrained format were next studied to establish the fractional helicity (fH) gained in constrained form (Figure 3). In doing so, it was found that no cassette in linear form exceeded 19% fH based on the raw 222 nm signal (Figure S5 and Table 1), whereas even the lowest lactamized peptide was 34% helical while displaying a characteristic α -helical signature (Table 1 and Figures 3 and S6). All cassettes exhibited helicity at 20 °C when lactamized, with further thermal scans undertaken to monitor the stability of the cassettes at 10 °C increments. At the final scan of 90 °C, although loss of stability for all lactamized cassettes was observed, all 16 presented an α -helical signature with an average of 38% helicity, indicating that

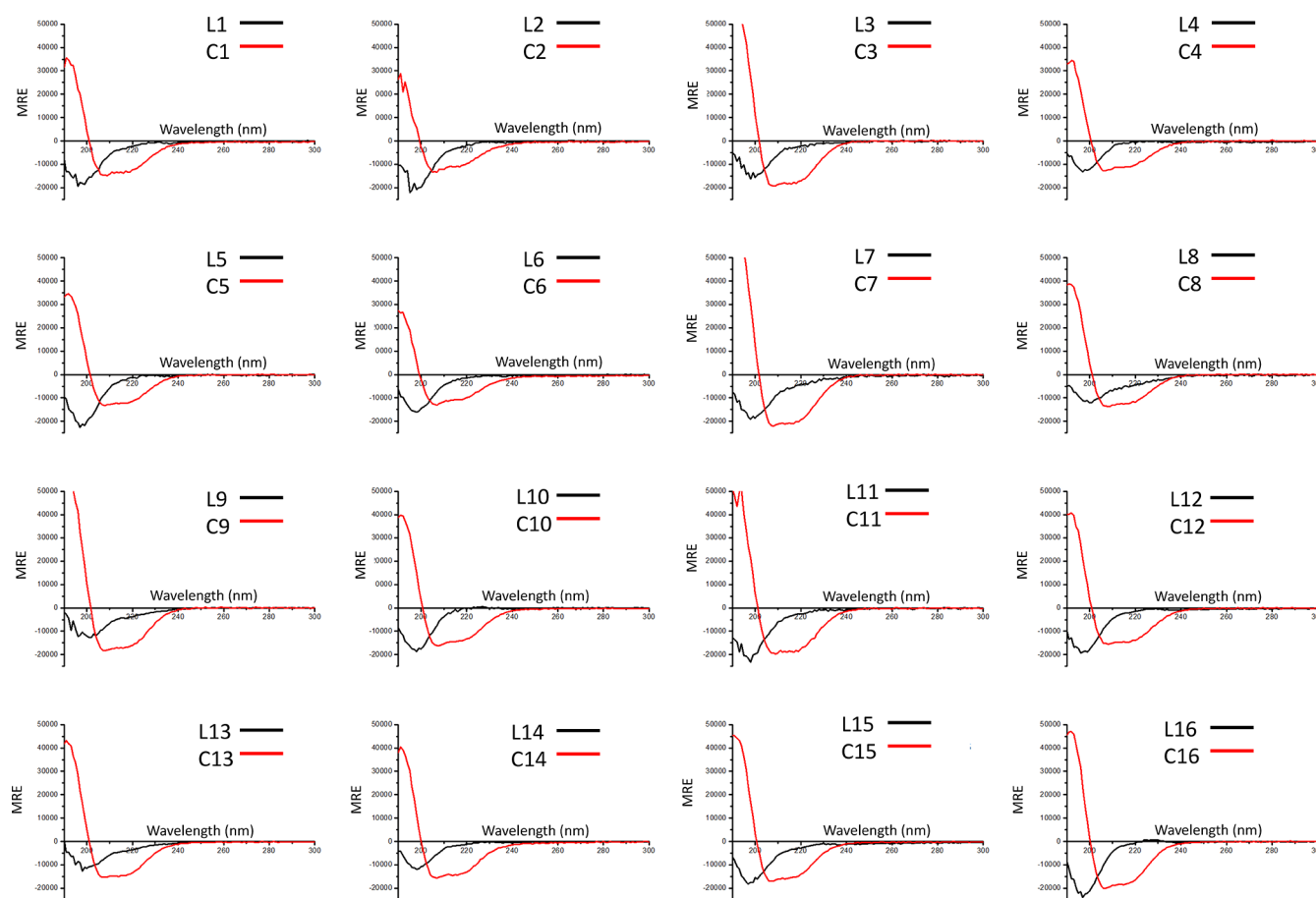


Figure 2. CD spectra for cassettes 1–16. These are shown in both the linear (black) and lactamized (red) forms. Spectra were measured at 20 °C at a total peptide concentration of 150 μ M and presented as mean residue ellipticity (MRE). All experiments were performed in 10 mM potassium phosphate and 100 mM potassium fluoride (pH 7). In all cases, the linear forms were characterized as having a random coil profile—with an average 222/208 nm ratio of 0.24. The lactamized cassettes displayed more of a helical profile—with an average ratio of the signal at 222/208 nm of 0.79.

helicity is maintained at higher temperatures where no structure would be expected for any linear counterpart. Constrained peptides displayed characteristic 222/208 nm minima, and consistent with lactamization, no cooperative unfolding profile was observed. Moreover, after heating to 90 °C, each cassette returned to within 5% of the original signal when incubated at 20 °C (Figure S6), indicating that any loss of structure is fully reversible.

Predicting the Stability of Cassettes with Heptads within cJun. To provide further evidence for the validity of our approach, in parallel to measuring individual cassette helicities, we used bCIPA to screen each of the 16 sequences against individual *gabcdef* heptads within a cJun coiled-coil target sequence. This was performed to provide a qualitative ranking of the most appropriate cassette to take forward when only considered as a linear heptad. bCIPA prediction is a useful as a tool to gain understanding of the most appropriate core and electrostatic options required for target stability as well as for specificity of interaction, for instance, in avoiding homodimerization over target interaction.^{20–22} Hence, bCIPA aids in selecting which cassettes are most appropriate for conjugation in targeting a coiled-coil sequence (i.e., which are predicted to bind *and* adopt a conformation of high helicity). These can then be taken forward for each given heptad within a target sequence. Because bCIPA is trained and validated on longer coiled-coil sequences, the values generated

are purely qualitative and are not treated as true T_m values. As shown in Table 2, the bCIPA values generated have been normalized to serve alongside helicity in predicting the most appropriate balance of helicity and affinity for each heptad within the target.

Combinational Design To Target cJun. Two four-heptad peptides were created to target the coiled-coil region of the oncogenic transcriptional regulator protein cJun. Using this approach, cassettes were selected that when conjugated in linear form were predicted to bind with high affinity to the cJun target sequence (Figure 4). In particular, two peptides were tested; one was predicted to contain cassettes that constrain poorly and therefore impart poor helicity upon the full-length peptide sequences (peptide 6–2–13–5; individual cassette fH values = 36, 34, 47, and 41%, respectively, Table 2, footnote c), and a second peptide, which was predicted to contain cassettes that constrain well and impart high levels of helicity upon the full-length peptide sequence (peptide 1–3–16–7; individual cassette fH values = 45, 60, 55, and 68%, respectively, Table 2, footnote b). Both peptides were selected by using a combination of (i) normalized bCIPA values and (ii) the fH of the lactamized cassettes (Table 2). Having calculated fH for all 16 lactamized cassettes, each was next individually screened against the first four heptads of cJun. This permitted selection from 16^4 (65 536) possible unique cassette arrangements to predict the most appropriate

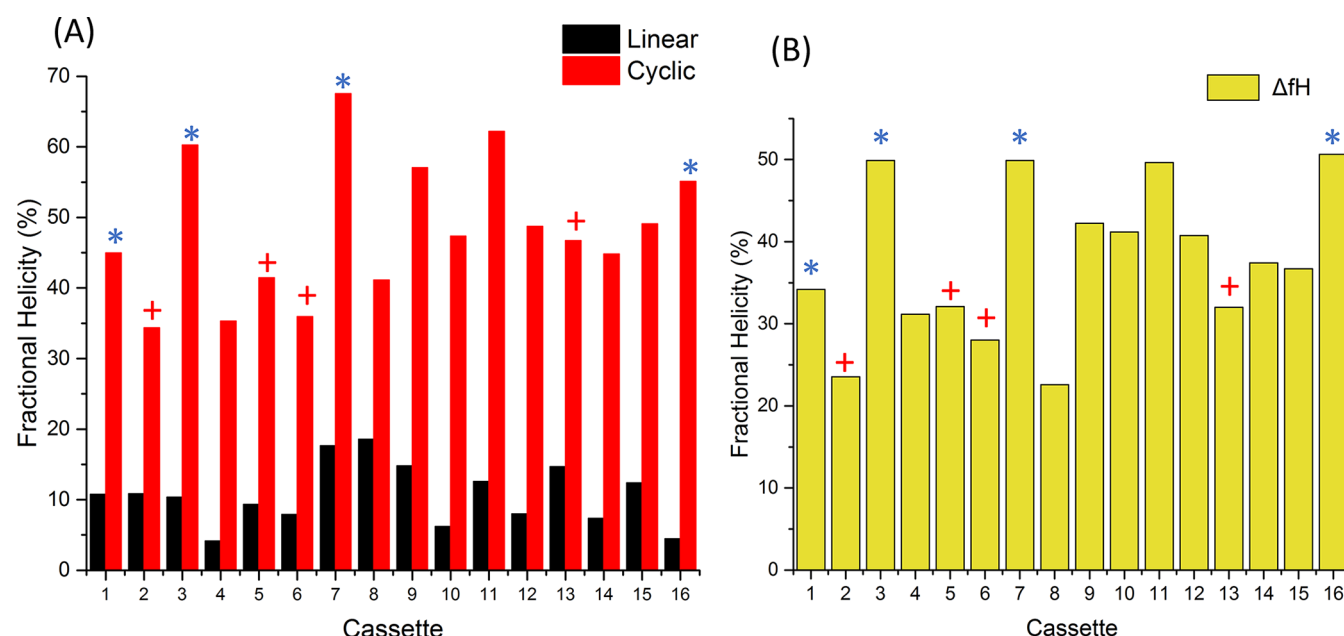


Figure 3. Fractional helicity (fH) data for cassettes 1–16. (A) shows values calculated from the 222 nm value from circular dichroism scans at 20 °C. For each cassette, the cyclic (lactamized) form has increased helicity compared to the linear (black), with an average increase of 37.6%. The ΔfH values (B) show that the largest increase in helicity was seen in cassette 16 (ΔfH = 50.6%), whereas the lowest increase in helicity was seen in cassette 8 (ΔfH = 22.6%). Highlighted are the cassettes used in the full-length peptide (*) and those used in the control (+).

Table 1. Comparative Helicity of Cassettes 1–16^a

Cassette Number	Sequence <i>gabcdef</i> <i>gaKALeD</i>	Linear fH (%)	Cyclic fH (%)	ΔfH (%)
1	(Y) <i>EIKALeD</i>	10.8	45.0	34.2
2	(Y) <i>KIKALeD</i>	10.8	34.4	23.5
3	(Y) <i>ELKALeD</i>	10.4	60.3	49.9
4	(Y) <i>KLKALeD</i>	4.1	35.3	31.2
5	(Y) <i>EVKALeD</i>	9.3	41.4	32.1
6	(Y) <i>KVKALeD</i>	7.9	35.9	28.0
7	(Y) <i>EIKALKD</i>	17.6	67.5	49.9
8	(Y) <i>KIKALKD</i>	18.6	41.1	22.6
9	(Y) <i>ELKALKD</i>	14.8	57.0	42.2
10	(Y) <i>KLKALKD</i>	6.2	47.4	41.2
11	(Y) <i>EVKALKD</i>	12.6	62.2	49.6
12	(Y) <i>KVKALKD</i>	8.0	48.7	40.7
13	(Y) <i>ENKALeD</i>	14.7	46.7	32.0
14	(Y) <i>KNKALeD</i>	7.4	44.8	37.4
15	(Y) <i>ENKALKD</i>	12.4	49.1	36.7
16	(Y) <i>KNKALKD</i>	4.5	55.1	50.6

^aSequences for cassettes 1–16 are shown with linear and cyclic fH values (and the difference) as determined by circular dichroism at 20 °C. Cassettes with the highest (red) and lowest (blue) cyclic fH values are highlighted.

sequence for effective binding. Combining cassettes 1, 3, 16, and 7 (Figure 4) led to the design of a constrained peptide predicted to form both a stable α -helix (average cassette helicity = 57%) but with the additional possibility of forming a higher affinity coiled-coil with the target cJun, relative to the linear counterpart (linear 1–3–16–7:cJun bCIPA predicted

$T_m = 38$ °C). As a control, cassettes were also conjugated, which were similarly highly ranked according to the bCIPA screening process. However, in contrast, these sequences were experimentally validated to be of low isolated helicity (average helicity = 39%, linear 6–2–13–5:cJun bCIPA predicted $T_m = 41$ °C). This resulted in a combination of cassettes 6, 2, 13,

Table 2. Selection of Cassettes for Conjugation^a

RIARLEE			KVKTLKA			QNYELAS			TANMLRE		
heptad 1			heptad 2			heptad 3			heptad 4		
cassette	heterodimer T_m (°C)	helicity (%)	cassette	heterodimer T_m (°C)	helicity (%)	cassette	heterodimer T_m (°C)	helicity (%)	cassette	heterodimer T_m (°C)	helicity (%)
2	31.1	34.4%	1	47.2	45.0%	1	12.5	45.0%	3	27.6	60.3%
6 ^c	29.8	35.9%	3 ^b	44.8	60.3%	7	10.7	67.5%	1	24.7	45.0%
4	28.7	35.3%	5	40.6	41.4%	3	10.2	60.3%	5 ^c	23.4	41.4%
1 ^b	25.8	45.0%	2 ^c	33.4	34.4%	13 ^c	8.7	46.7%	9	23.4	57.0%
5	24.5	41.4%	7	33.4	67.5%	9	8.4	57.0%	7 ^b	20.5	67.5%
3	23.4	60.3%	4	31.0	35.3%	2	8.4	34.4%	11	19.2	62.2%
8	19.8	41.1%	9	31.0	57.0%	15	6.9	49.1%	4	16.2	35.3%
12	18.5	48.7%	13	27.5	46.7%	8	6.6	41.1%	13	15.5	46.7%
10	17.4	47.4%	11	26.9	62.2%	4	6.0	35.3%	2	13.3	34.4%
14	16.7	44.8%	6	26.9	35.9%	5	6.0	41.4%	10	12.0	47.4%
7	14.4	67.5%	8	19.7	41.1%	14	4.5	44.8%	6	12.0	35.9%
11	13.1	62.2%	10	17.3	47.4%	10	4.2	47.4%	15	11.3	49.1%
9	12.0	57.0%	14	13.7	44.8%	11	4.2	62.2%	8	9.1	41.1%
13	11.3	46.7%	15	13.7	49.1%	16 ^b	2.7	55.1%	12	7.8	48.7%
16	5.4	55.1%	12	13.1	48.7%	6	1.8	35.9%	14	4.2	44.8%
15	0.0	49.1%	16	0.0	55.1%	12	0.0	48.7%	16	0.0	55.1%

^aHeptads 1–4 of cJun (listed in top row) have been predicted against each of the cassettes, alongside the helicity of the lactamized forms from circular dichroism. The bCIPA heterodimer values (cassette–cJun) have been normalized to account for all values <0. For heptad 3, the only cassettes considered were 13–16 (since they contain Asn at the a position). Cassettes were chosen sequentially, and no cassette was selected multiple times. In total, 16⁴ cassette combinations were possible leading to 65 536 potential peptide sequences. ^bCassettes chosen for the full-length alpha helix. ^cCassettes chosen for the full-length control peptide, chosen by the predicted T_m alone.

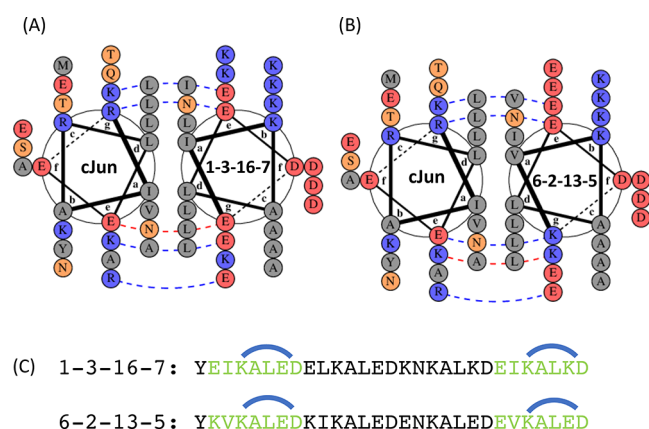


Figure 4. Design of lactamized sequences. Shown are 1–3–16–7 (A) and control 6–2–13–5 (B). In both cases, cassettes were chosen independently of one another. For 1–3–16–7, considering the cassettes in isolation, there is a mixture of electrostatically repulsive (cassette 1), favorable (cassette 3), and nonoptimal (cassette 16). For heptad 3 (C), the Asn–Asn interaction has been selected for at the core. For 6–2–13–5, there is a mixture of electrostatically favorable (cassette 6), partially favorable (cassettes 2, 5), and nonoptimal (cassette 13). For heptad 3 (C), the Asn–Asn interaction has been selected for at the core. (C) Sequences for peptides 1–3–16–7 and 6–2–13–5. The addition of a Tyr at the N-terminus creates a 29 residue peptide that starts at the f position and ends at the f position. Highlighted in green are the cassettes that have Lys/Asp lactamization (1 and 7/6 and 5).

and 5 (Figure 4). This permitted the comparison of two sequences, both of which were predicted to engage with the target as linear 28-mers (Figure 4). Computationally, both peptides score very highly, with 6–2–13–5 computationally ranked as #145 and 1–3–16–7 as #450—both within the top 1% of the library of full-length peptides. Indeed, both

sequences contain the same net favorable electrostatic contribution (see below) and comparable core contributions as well as Asn residues at position a3 to promote asymmetric side-chain–side-chain hydrogen bonding with the corresponding Asn in the cJun partner strand.^{26,27} However, while 1–3–16–7 was expected to lactamase well and translate into improved binding, 6–2–13–5 was expected to lactamase poorly leading to a lower gain of coiled-coil stability. During the design process, eight unique cassettes were selected to avoid duplication of any modules and to widen the potential understanding of each cassette use in combination. Both sequences, 1–3–16–7 (YEIKALEDELKALEDKNKALKDEIKALKD) and 6–2–13–5 (YKVKALEDKIKALEDENKALEDEVKALED), were synthesized with a Tyr at the N-terminus for quantification by UV absorbance. Both sequences terminate at the f-position of the fourth heptad (Figure 4) and contained two lactam bridges—one at either termini. As helicity is thought to propagate from the N-terminus to the C-terminus,³ stabilization via heptads 1 and 4 was predicted to promote overall helicity toward heptads 2 and 3 at the helix interior. In contrast to previous work,^{18,21,22,28} the 29-mers were not capped with helicity promoting residues (AS at N-terminus and GAP at the C-terminus) to further probe the effect of lactamization in directly promoting helicity.

Upon inspection of the helical wheels (Figure 4), a pattern of electrostatically favorable and unfavorable interactions was observed. As selected by bCIPA, as a potential homodimer peptide 1–3–16–7 displayed six repulsive electrostatic interactions (four Glu–Glu and two Lys–Lys) and only two attractive interactions (two Glu–Lys). In the 6–2–13–5 control, four repulsive and four attractive interactions are present. As potential heterodimers with cJun, both peptides contain four favorable and one unfavorable interaction. As discussed previously,²² cJun contains Gln and Ala at the g3 and e3 positions. The a3–a3' was chosen to generate an Asn–Asn

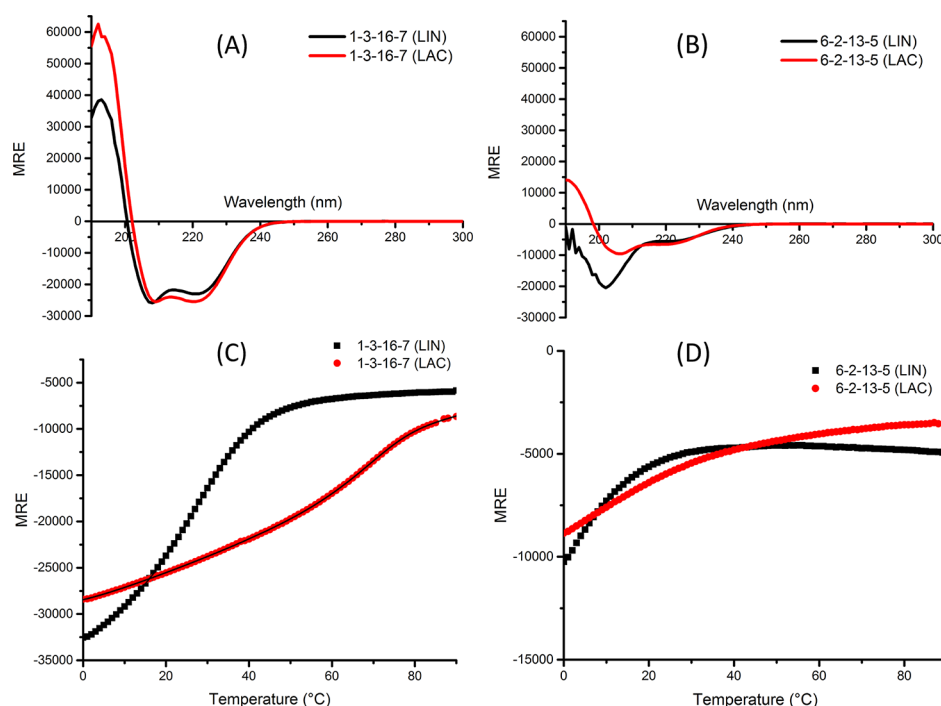


Figure 5. CD spectra and thermal denaturation data for linear and lactamized peptides in isolation. Shown are data for 1–3–16–7 (A,C) and 6–2–13–5 (B,D). Spectra were measured at 20 °C at a total peptide concentration of 150 μ M and presented as mean residue ellipticity (MRE). The minima at 208 and 222 nm are indicative of a helical structure, with the 222/208 nm ratio of the lactamized 1–3–16–7 showing more structure (222/208 nm = 1.01) than the linear (222/208 nm = 0.89). The lactamized 6–2–13–5 shows increased helical structure (222/208 nm = 0.69) compared to the linear (222/208 nm = 0.44). This suggests that the addition of lactam bridges improves the α -helicity in both peptides. Thermal denaturation profiles of linear and lactamized 1–3–16–7 (C) and 6–2–13–5 (D) peptides were taken using 1 °C increments and tracking the 222 nm signal at 150 μ M. Lactamized 1–3–16–7 shows an increase in the transition midpoint with a T_m of 67.0 °C compared to the linear T_m of 29.1 °C. Lactamized 6–2–13–5 demonstrates a change* in the transition midpoint when in complex with cJun compared to the linear*. All experiments were performed in 10 mM potassium phosphate and 100 mM potassium fluoride (pH 7). Where possible (C), data were fitted to the two-state model.

interaction for both sequences because of the favorable effect of this interaction on specificity and on the oligomeric state.

Structural Stability of Terminally Lactamized Peptides. An analysis of the global secondary structure of the full-length linear and lactamized peptides was conducted, both in isolation and in complex with cJun. CD spectra showed that 1–3–16–7 displayed high levels of helicity in both linear (65.3%) and lactamized (71.4%) forms (Figure 5a). At 20 °C, the stability of the two are comparable ($\Delta fH = 6.1\%$), with the lactamized version displaying an improved helical signature (222/208 nm ratio = 1.01). In contrast, the linear 6–2–13–5 peptide displays comparatively low helicity (Figure 5b, $fH = 18.2\%$), increasing upon lactamization to display an improved overall α -helical signature (222/208 nm = 0.69) but with little change in overall helicity ($fH = 20.4\%$).

Global secondary structure and stability was next monitored by incubating peptides with cJun. At 20 °C, the linear form of 1–3–16–7 (Figure 6a) showed a helical profile (222/208 nm = 0.87) with a fractional helicity greater than the average ($fH = 55.2\%$) of the component peptides ($\Delta fH = 8\%$). For the lactamized version (Figure 6b), the profile remained highly helical (222/208 nm = 0.90) with helicity ($fH = 52.0\%$) negligibly greater than the average of the two component peptides ($\Delta fH = 1.75\%$). In both the linear and lactamized forms of 6–2–13–5 (Figure 7a,b), levels of helicity were lower. The linear form incubated with cJun displayed no increase in signal ($\Delta fH = 0.1\%$) relative to the component

peptides. Similarly, the lactamized form showed a level of helicity ($fH = 25\%$) similar to that of the average.

Thermal Denaturation Profiles. Having observed varying levels of helicity between all peptides, CD thermal denaturation experiments were next performed to establish the extent to which the stability of the complexes changed upon introduction of the constraints. In these experiments, the 222 nm signal was monitored at 1 °C increments from 0 to 90 °C. For linear 1–3–16–7 in isolation (Figure 5C), an interaction was observed ($T_m = 29.1$ °C), which upon lactamization led to an increase in thermal stability ($T_m = 67.0$ °C). Moreover, the signal persisted even at high temperatures, implying that the lactamized form retains residual thermal stability relative to the linear form. When in complex with cJun (Figure 6C,D), the linear peptide exhibited an increase in T_m (34 °C), with the lactamized form displaying a marked increase in thermal stability over the average of the component peptides, demonstrating that the lactamized form of the peptide is able to preferentially bind cJun. The ΔT_m (1–3–16–7_{LAC}:cJun – 1–3–16–7_{LIN}:cJun) was found to be 23 °C, demonstrating that dual lactamization is tolerated and translates into increase target affinity.

For 6–2–13–5, both peptides displayed lower levels of stability (Figure 7C,D), with T_m values unable to be derived from the thermal melt profiles. Consistent with CD spectra, this implies a much lower level of thermal stability in isolation. However, upon incubation with cJun, the linear form displays higher levels of heterodimeric thermal stability ($T_m = 18.9$ °C),

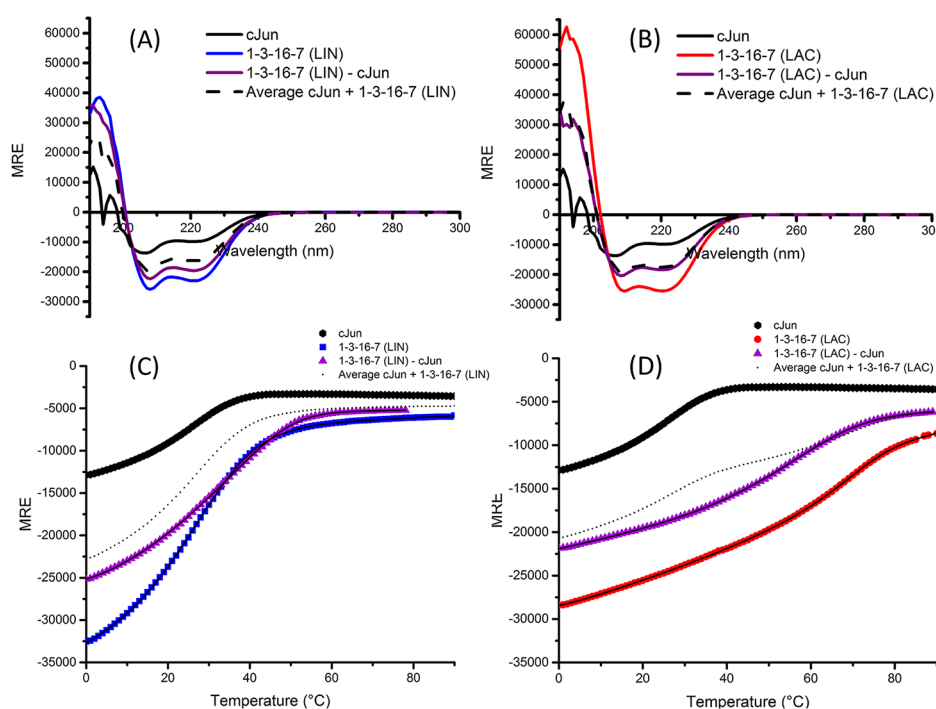


Figure 6. CD spectra and thermal denaturation data for linear and lactamized 1–3–16–7 with cJun. Shown are data for the 1–3–16–7 peptide in the linear (A,C) and lactamized (B,D) forms in complex with cJun. Spectra were measured at 20 °C at a total peptide concentration of 150 μ M and presented as mean residue ellipticity (MRE). The minima at 208 and 222 nm are indicative of a helical structure, with the 222/208 nm ratio of the linear peptide with cJun showing a similar structure (222/208 nm = 0.87) to the homomeric complex (222/208 nm = 0.89). The lactamized form displays a decreased helical structure (222/208 nm = 0.90) compared to the homomeric state (222/208 nm = 1.01). Thermal denaturation profiles of homomeric 1–3–16–7 and that in complex with cJun (C,D) were measured using 1 °C increments and tracking the 222 nm signal at 150 μ M. Linear 1–3–16–7 displays an increase in the transition midpoint when in complex with cJun (C), with a T_m of 34 °C compared to the homomer (T_m = 29.1 °C). Lactamized 1–3–16–7 demonstrates a decrease in the transition midpoint, with a T_m value of 57.2 °C when in complex with cJun compared to the homomer T_m of 67 °C (cJun T_m = 25.8 °C). All experiments were performed in 10 mM potassium phosphate and 100 mM potassium fluoride (pH 7). Where possible (D), data were fitted to the two-state model.

which is increased upon lactamization (T_m = 29.2 °C). The ΔT_m (6–2–13–S_{LAC}:cJun – 6–2–13–S_{LIN}:cJun) was found to be 10 °C. Overall this suggests that, as reflected in the lower levels of helicity relative to 1–3–16–7_{LAC}:cJun (fH = 25 vs 52%), the constraint is less well tolerated. Moreover, this consequently translates into a lower overall improvement in target affinity relative to the linear counterpart. Figure 8 highlights the relevance of these modular designs in the wider scheme of engineered peptides that can tolerate constraints and that lead to a demonstrable increase in target affinity for a defined coiled-coil target. As previously discussed,²² there are indirect parallels that can be drawn based on T_m . Previous isothermal calorimetry work characterizing the biophysical properties of peptides demonstrated a FosW–cJun interaction K_D value of 39 nM,²⁹ which displayed a thermal stability similar to that of the lactamized 1–3–16–7 (T_m = 63 °C). Similarly, the 1–3–16–7_{LAC}:cJun complex reported the same helicity (fH = 52%) and a similar T_m and K_D to that of cFos-24 (T_m = 58 °C; K_D of 7.25 μ M).³⁰ The comparative values provide a broad understanding of enhanced structural stability displayed in lactamized complexes and demonstrate the potential impact lactamization may have on biophysical properties and the ability to predict such properties based upon modular design and experimental validation.

Helix Nucleation. The folding of the α -helix is due in large part to the stabilization conferred by sequential $i \rightarrow i + 4$ hydrogen bonds. Moreover, α -helices display an overall dipole moment with a pair of terminal microdipoles because of a lack

of intrahelical hydrogen bonds at the termini. In particular, unsatisfied hydrogen bonding by the first four >N–H groups at the N-terminus can lead to a partial positive charge (δ^+). Similarly the last four >C=O groups at the C-terminus can lead to a partial negative charge (δ^-).³¹ Instead, these groups are often capped by alternative hydrogen bond partners that are provided by helix-capping motifs.³² It is therefore generally accepted that charge–macro-dipole interactions can play a small role in enhancing the stability of a helix—with negatively charged residues at the N-terminus and positively charged residues at the C-terminus that can counter the effects of the dipole. As highlighted in Figure 4, the designs of our full-length α -helix could be optimized to counter the dipole. For peptide 1–3–16–7, Glu was introduced at the N-terminus and Lys at the C-Terminus. However, the control peptide (6–2–13–5) did not provide charge complementarity for the helix macrodipole, containing a Lys at the N-terminus and Glu at the C-terminus. The former peptide is more stable than the latter (either as a homomer or in complex with cJun). Coupled with constraints that are less well tolerated in individual cassettes for the control sequence (average helicity of component cassettes = 57% vs 38%), this suggests that charge stabilization at both termini, even with addition of less well tolerated constraints, could be another influencing factor. Coupling the structural stability lactam bridges confer with favorable terminal residues may therefore represent another way to introduce additional target affinity into the peptide.

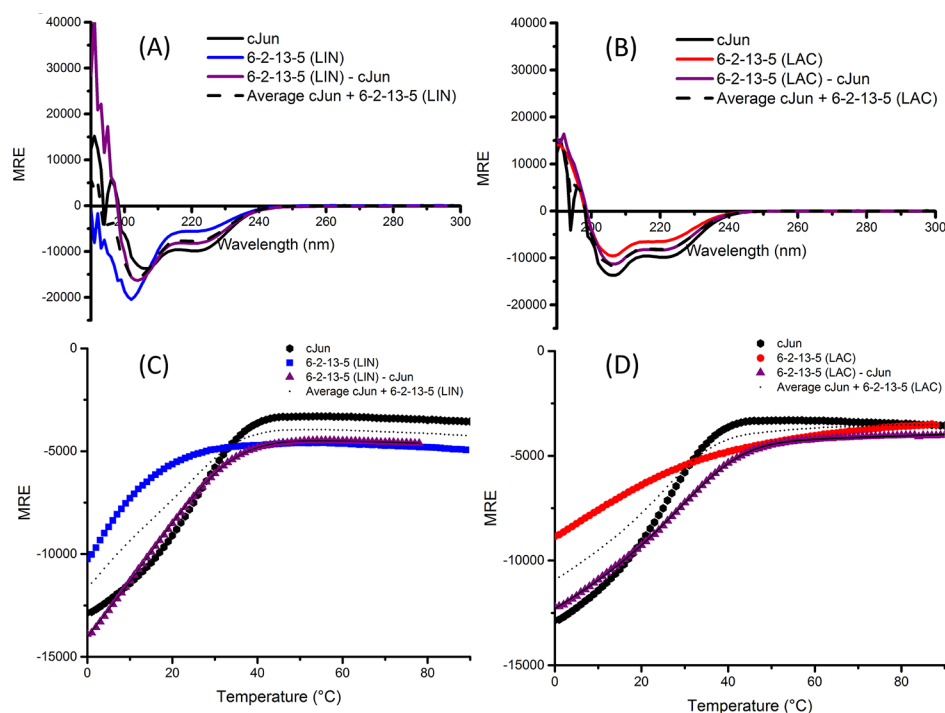


Figure 7. CD spectra and thermal denaturation data for linear and lactamized 6–2–13–5 with cJun. Shown are data for the 6–2–13–5 peptide in the linear (A,C) and lactamized (B,D) forms in complex with cJun. Spectra were measured at 20 °C at a total peptide concentration of 150 μ M and presented as mean residue ellipticity (MRE). The minima at 208 and 222 nm are indicative of various levels of helical structure, with the 222/208 nm ratio of the linear peptide with cJun showing increased structure (222/208 nm = 0.60) to the homomeric complex (222/208 nm = 0.44). The lactamized form shows increased structure (222/208 nm = 0.74) compared to the homomeric state (222/208 nm = 0.69). This implies that the addition of cJun increased the helicity of the lactamized and linear peptide. Thermal denaturation profiles of homomeric 6–2–13–5 and that in complex with cJun (C,D) were taken using 1 °C increments and tracking the 222 nm signal at 150 μ M. Linear 6–2–13–5 shows an increase in the transition midpoint when in complex with cJun (C), demonstrating a T_m of 18.9 °C compared to the homomer*. Lactamized 6–2–13–5 demonstrates an increase in the transition midpoint, with a T_m value of 29 °C when in complex with cJun compared to the homomer*. All experiments were performed in 10 mM potassium phosphate and 100 mM potassium fluoride (pH 7). Where possible (D), data were fitted to the two-state model. *Denaturation profiles for homomeric 6–2–13–5 (linear and lactamized) were unable to be fit.

Sequence-Specific Helicities. As a result of the limited randomization and size of the cassette library, there exists an opportunity to further study the effects of the different sequence combinations from the perspective of helical stability. Despite focusing on peptides longer than 10 residues, the idea of sequence-specific stability (in the context of positions *g/e*) has been discussed previously.²⁰ The presence of N-terminal Glu and C-terminal Lys might have been expected to lead to the highest helicity values with every different core options. As is shown in Figure 9, there does appear to be a pattern within the library. For each of the four core options, it would be expected that the Glu/Lys selection (with a positively charged C-terminus) would yield stability similar to that of the peptide 1–3–16–7. This is partially correct as, regardless of the core arrangement, the E/K cassettes display the highest (Ile/Val) or second highest (Leu/Asn) helicities. When observing the Ile/Leu/Val library members, this electrostatic arrangement resulted in cassettes with helicities 10–23% higher than the next most stable. In the case of Asn, the helicity values are far more similar, with a difference of 10% between the four. In this case, the polarity of the core Asn appears likely to be influencing the effect that these terminal charged residues usually impart upon the macrodipole and therefore upon helicity. When considering the inverse, the lowest helicity cassettes are those with Lys at the N-terminus and Glu at the C-terminus. This is logical when considering the arguments

proposed previously regarding the helix macrodipole within the 29-mer.

Conclusion. We have demonstrated that modular *gabcdef* heptad cassettes can be lactamized via side-chain to side-chain $K \rightarrow D$ ($i \rightarrow i + 4$) bridges as a mechanism to impart high helicity upon otherwise unstructured sequences. We have shown that although all of the sequences adopt helical conformations, the fractional helicity observed varies in magnitude from 34 to 68% across the 16 lactamized cassettes. It was therefore possible to use the cassettes to create a library that can be conjugated into four repeats ($16^4 = 65\,536$ possible combinations) toward creating coiled-coil antagonists. Furthermore, we utilized the bCIPA library screening algorithm (<http://people.bath.ac.uk/jm2219/biology/iscan.zip>) as a means to select peptides that (i) best engage with a heptad counterpart within a target helix and (ii) have a low tendency to self-associate.

Conjugating individual cassettes of high measured helicity (1, 3, 16, 7 = 45, 60, 55, 68% respectively) into a longer sequence constrained by terminal lactams was shown to impart further helicity (fH = 65% linear vs 71% lactamized), which consequently translated into an increased T_m with the cJun target (34 °C linear vs 57 °C lactamized). In contrast, a control sequence in which component cassettes were known to lactamized comparatively poorly (6, 2, 13, 5 = 36, 34, 47, 41%), was lactamized at the termini with only a negligible gain in helicity observed (fH = 18% linear vs 20% lactamized),

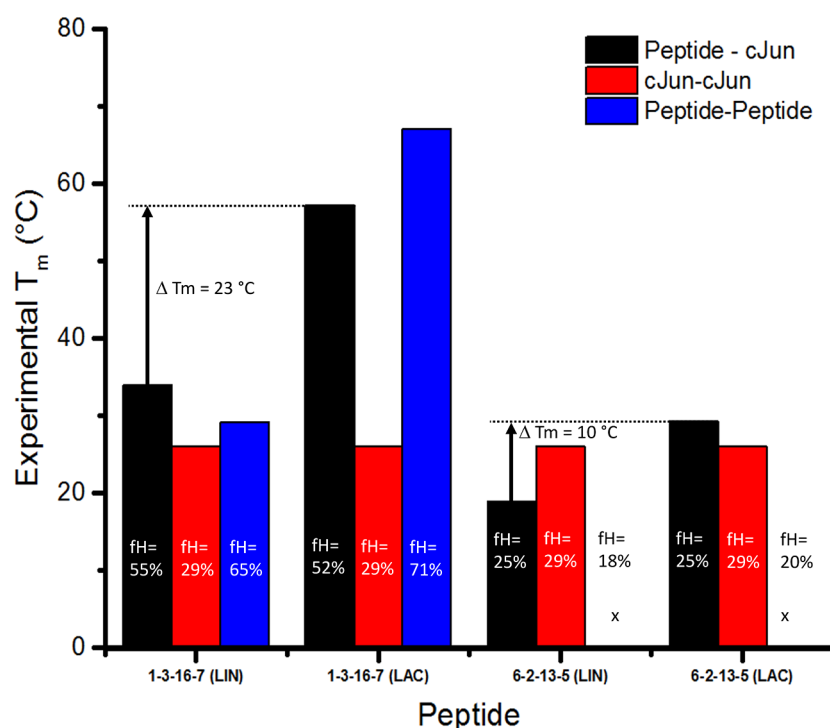


Figure 8. Comparison of the full-length peptides. The T_m of lactamized 1–3–16–7 was 67 °C as a homomer and 57.2 °C in complex with cJun. x indicates an inability to determine a T_m from the melt profile and represents a low stability complex.

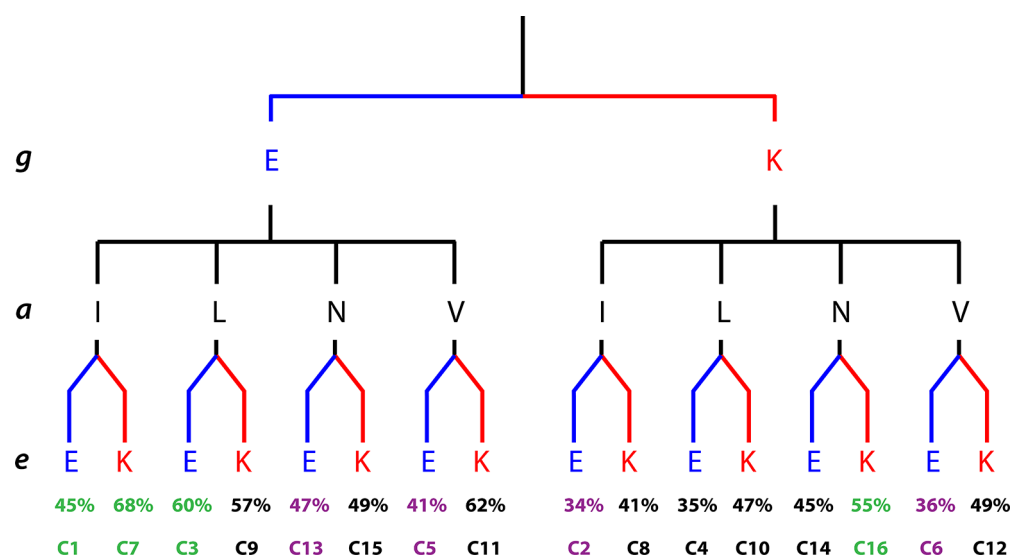


Figure 9. Analysis of individual cassette composition demonstrates stability to be sequence-specific. Across the library, a Lys/Glu combination at positions g/e results in decreased helicity (e.g., cassettes 2/4/6). Conversely, the selection of Glu/Lys at these positions results in higher levels of helicity—with a delta of 10–23% between cassettes 7/9/11 and the next most stable cassettes. Asn containing cassettes exhibited similar patterns, although the differences between helicity values within the four cassettes was shown to be minimal (a delta of 10% within C13–16).

which translated into only a modest gain in T_m with the cJun target (19 °C linear vs 29 °C lactamized).

Because all cassettes exhibited increased helicity when lactamized, this poses an interesting question regarding the ability of certain sequences to adopt a more helical conformation than others (Figures 3 and 9). Because the possible permutations within the 16 peptides were limited, there is a limit to the information that can be drawn in terms of electrostatic and core configuration that optimize the stability of the cassettes. However, the placement of Glu at the N-terminal g position appears to impart increased helicity over all

other options unless Asn resides at the core a position. The placement of Lys at the C-terminal e position appears to be less influential on its own but does complement the placement of an N-terminal Glu at the g position. Placement of Lys at the g position and Glu at the e position results in the lowest helical values of all cassettes but is tempered slightly by the placement of Asn at the a position. With more data, this presents an opportunity to optimize computational techniques employed for prediction. As previously shown,²² bCIPA is capable of being trained on a specific subset of coiled-coil forming sequences. With further experimental exploration of the

cassette library, reflecting other cassette types observed within natural coiled coils, it will be possible to extract more helical predictors toward developing a more robust prediction algorithm that can specifically consider lactamized peptides. For instance, further interrogation of the system could involve expanding the cassette library to include residues with larger side-chains or side-chains which have been previously shown to be nonoptimal at certain sequence positions in the design of α -helical peptides. In this study, we have probed the ability of given coiled-coil sequences to tolerate the introduction of K \rightarrow D lactam bridges between solvent exposed *b* and *f* residues. Once established that lactams are tolerated, then the precise level of helical induction can be calculated. However, the assumption that molecules with increasingly higher helicity might better form coiled coils is an oversimplification.^{29,30,33} Rather, there is likely a limit to the entropy value of preorganizing a helical structure. Above a certain threshold helicity, conformational entropy may serve to oppose coiled-coil formation,³⁴ perhaps reflecting the need for some residual helix flexibility to enable the distortion necessary for supercoiling.^{35,36}

The properties of cassettes with high levels of stability in comparison to others will allow further understanding of the rules, by which *b* \rightarrow *f* lactamization imparts helicity. As further experimental data is acquired, computational techniques employed in the selection of cassettes for specific functions will be refined. These “off-the-shelf” sequences can be combined with computational screening and applied to a wide range of parallel dimeric coiled-coil dimerization domains as a generalized method to ablate or even agonize the function of many different PPIs in which coiled coils are found.

MATERIALS AND METHODS

bCIPA Peptide Library Screening. bCIPA screening was performed as described previously.²² Briefly, individual sequences of cassettes were calculated as homodimers and with the cJun target sequence using software based on the bCIPA algorithm.²⁸ Thermal denaturation (T_m) values were normalized. Rink amide ChemMatrix resin was obtained from PCAS Biomatrix, Inc. (St.-Jean-sur-Richelieu, Canada); Fmoc L-amino acids and 2-(1*H*-benzotriazole-1-yl)-1,1,3,3-tetra-methyluronium hexafluorophosphate (HBTU) or benzotriazol-1-yl-ox-ytripyrrolidinophosphonium hexafluorophosphate (PyBOP) were obtained from AGTC Bioproducts (Hessle, UK). All other reagents were of peptide synthesis grade and obtained from Thermo Fisher Scientific (Loughborough, UK). Peptides were synthesized on a 0.1 mmol scale on a PCAS ChemMatrix Rink amide resin using a Liberty Blue microwave peptide synthesizer (CEM; Matthews, NC) employing Fmoc solid-phase techniques³⁷ with repeated steps of coupling, deprotection, and washing (4 \times 5 mL dimethylformamide). Coupling was performed as follows: Fmoc amino acid (5 equiv), HBTU or PyBOP (4.5 equiv), and diisopropylethylamine (10 equiv) in dimethylformamide (5 mL) for 5 min with 35 W microwave irradiation at 90 °C. Deprotection was performed as follows: 20% piperidine in dimethylformamide for 5 min with 30 W microwave irradiation at 80 °C. Following synthesis, peptides were acetylated using acetic anhydride (3 equiv) and diisopropylethylamine (4.5 equiv) in dimethylformamide (2.63 mL) for 20 min. Deprotection of acid labile Asp(oPip) and Lys(Mtt) side-chain-protecting groups was achieved by repeated washing of the resin in dichloromethane, followed by repeated washes in dichloromethane (2% TFA), dichloromethane, and finally dimethylformamide. Resin was next incubated for 7 h at 55 °C in 2-(1*H*-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (1 mL), diisopropylethylamine (1 mL), and dimethylformamide (3 mL). Resin was filtered and cleaved from the resin with concomitant removal of side-chain-protecting groups by treatment with a cleavage mixture (10 mL)

consisting of TFA (95%), triisopropylsilane (2.5%), and H₂O (2.5%) for 4 h at RT. Suspended resin was removed by filtration, and the peptide was precipitated using three rounds of crashing in ice-cold diethyl ether, vortexing, and centrifuging. The pellet was then dissolved in 1:1 MeCN/H₂O and freeze-dried. Purification was performed by RP-HPLC using a Phenomenex Jupiter Proteo (C18) reverse-phase column (4 μ m, 90 Å, 10 mm inner diameter \times 250 mm long). Eluents used were as follows: 0.1% TFA in H₂O (a) and 0.1% TFA in ACN (b). The peptide was eluted by applying a linear gradient (at 3.5 mL/min) of 5–95% B over 50 min. Fractions collected were examined by electrospray MS, and those found to contain exclusively the desired product were pooled and lyophilized. Analysis of the purified final product by RP-HPLC indicated a purity of >95%.

Peptide Quantification. Peptide concentrations were determined in ddH₂O or CD buffer (10 mM potassium phosphate and 100 mM potassium fluoride, pH 7) against the appropriate blank using the 280 nm absorbance maxima of the Tyr residue within each peptide (1209 M⁻¹ cm⁻¹). Prior to each measurement, samples were centrifuged at 13 000 rpm for 2 min to ensure that only soluble peptide was quantified. During this step, no precipitate was observed, indicating that all peptides displayed high levels of solubility. Measurements were taken using a Varian Cary 50 Conc UV Spectrophotometer using a 1 cm path length quartz cell.

Circular Dichroism. CD was carried out using an Applied Photophysics Chirascan CD Apparatus (Leatherhead, UK) using a 200 μ L sample in a CD cell with a 1 mm path length. Samples contained 150 μ M total peptide (Pt) concentration at equimolar concentration for heterodimeric solutions (i.e., 75 μ M per peptide) and suspended in 10 mM potassium phosphate and 100 mM potassium fluoride at pH 7 for 30 min prior to analysis. The CD spectra of samples were scanned between 300 and 190 nm in 1 nm steps, averaging 0.5 s at each wavelength. Three scans at 20 °C were averaged to assess helical levels and CC structure. For thermal scans, spectra were scanned as above in 10 °C steps from 20 to 90 °C. Each temperature point was held for 1 min to equilibrate sample before scanning. Raw data (ellipticities) were collected and averaged, and data were converted to molar residue ellipticities (MRE).

All spectral data was converted to fractional helicity (fH) values according to the equation

$$\begin{aligned} \text{fH} &= \frac{\theta_{222} - \theta_c}{\theta_{222\infty} - \theta_c} \\ \theta_c &= 2220 - (53 \times T) \\ \theta_{222\infty} &= (-44\,000 + (250 \times k)) \times \left(1 - \frac{k}{\text{Nr}}\right) \end{aligned} \quad (1)$$

where the wavelength-dependent constant $k = 2.4$ (at 222 nm), Nr = number of residues, and T = temperature (°C).

Thermal Denaturation. Thermal denaturation experiments were performed at 150 μ M in a buffer of 10 mM potassium phosphate and 100 mM potassium fluoride at pH 7 using an Applied Photophysics Chirascan Circular Dichroism Spectrometer. For all thermal denaturation experiments involving longer peptides, a stepping gradient was applied from 0 to 90 °C using 1 °C increments. Each temperature point was held for 30 s to equilibrate the sample before measuring ellipticity at 222 nm. Melting profiles were converted to equilibrium denaturation curves and fitted using a two-state model, derived via modification of the Gibbs–Helmholtz equation to yield the melting temperature (T_m).²⁸

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acscchembio.9b00265.

Table S1. Electrospray mass spectrometry data of linear and lactamized peptides; Table S2. Helical and thermal stability of linear and lactamized peptides; Figure S1. Electrospray mass spectrometry data from the purified cassettes 1–4; Figure S2. Electrospray mass spectrometry data from the purified cassettes 5–8; Figure S3. Electrospray mass spectrometry data from the purified cassettes 9–12; Figure S4. Electrospray mass spectrometry data from the purified cassettes 13–16; Figure S5. Thermal spectra for cassettes 1–16 in linear form; Figure S6. Thermal spectra for cassettes 1–16 in lactamized form; Figure S7. Electrospray mass spectrometry data from the purified full-length 1–3–16–7; Figure S8. Electrospray mass spectrometry data from the purified full-length 6–2–13–5 (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: j.mason@bath.ac.uk. (J.M.M.)

ORCID

Jody M. Mason: 0000-0002-4118-1958

Author Contributions

A.L. conducted experiments and synthesized, purified, and characterized peptides and cJun. J.M.M. directed the research. Both authors participated in experimental design, analysis of the data, and writing of the paper.

Notes

The authors declare the following competing financial interest(s): J.M.M. is an advisor to Sapience Therapeutics. A.L. has no financial or commercial conflict to declare.

ACKNOWLEDGMENTS

J.M. and A.L. thank the University of Bath for a Studentship. J.M. is also grateful to Cancer Research UK (A11738 and A26941) and to the BBSRC (BB/R017956/1) and EPSRC (EP/M001873/1).

REFERENCES

- (1) Walensky, L. D., and Bird, G. H. (2014) Hydrocarbon-Stapled Peptides: Principles, Practice, and Progress. *J. Med. Chem.* 57, 6275–6288.
- (2) De Araujo, A. D., Hoang, H. N., Kok, W. M., Diness, F., Gupta, P., Hill, T. A., Driver, R. W., Price, D. A., Liras, S., and Fairlie, D. P. (2014) Comparative α -helicity of cyclic pentapeptides in water. *Angew. Chem., Int. Ed.* 53, 6965–6969.
- (3) Acharyya, A., Ge, Y., Wu, H., DeGrado, W. F., Voelz, V. A., and Gai, F. (2019) Exposing the Nucleation Site in α -Helix Folding: A Joint Experimental and Simulation Study. *J. Phys. Chem. B* 123, 1797.
- (4) Verdine, G. L., and Hilinski, G. J. (2012) Stapled Peptides for Intracellular Drug Targets. *Methods Enzymol.* 503, 3–33.
- (5) Tan, Y. S., Lane, D. P., and Verma, C. S. (2016) Stapled peptide design: principles and roles of computation. *Drug Discovery Today* 21, 1642–1653.
- (6) Pal, S., and Prabhakaran, E. N. (2018) Hydrogen bond surrogate stabilized water soluble 310-helix from a disordered pentapeptide containing coded α -amino acids. *Tetrahedron Lett.* 59, 2515–2519.
- (7) Patgiri, A., Jochim, A. L., and Arora, P. S. (2008) A Hydrogen Bond Surrogate Approach for Stabilization of Short Peptide Sequences in α -Helical Conformation. *Acc. Chem. Res.* 41, 1289–1300.
- (8) Estieu-Gionnet, K., and Guichard, G. (2011) Stabilized helical peptides: overview of the technologies and therapeutic promises. *Expert Opin. Drug Discovery* 6, 937–963.
- (9) Lau, Y. H., De Andrade, P., Wu, Y., and Spring, D. R. (2015) Peptide stapling techniques based on different macrocyclisation chemistries. *Chem. Soc. Rev.* 44, 91–102.
- (10) Pelay-Gimeno, M., Glas, A., Koch, O., and Grossmann, T. N. (2015) Structure-Based Design of Inhibitors of Protein-Protein Interactions: Mimicking Peptide Binding Epitopes. *Angew. Chem., Int. Ed.* 54, 8896–8927.
- (11) Shepherd, N. E., Hoang, H. N., Abbenante, G., and Fairlie, D. P. (2005) Single turn peptide α helices with exceptional stability in water. *J. Am. Chem. Soc.* 127, 2974–2983.
- (12) Shepherd, N. E., Abbenante, G., and Fairlie, D. P. (2004) Consecutive cyclic pentapeptide modules form short α -helices that are very stable to water and denaturants. *Angew. Chem., Int. Ed.* 43, 2687–2690.
- (13) Tala, S. R., Schnell, S. M., and Haskell-Luevano, C. (2015) Microwave-assisted solid-phase synthesis of side-chain to side-chain lactam-bridge cyclic peptides. *Bioorg. Med. Chem. Lett.* 25, 5708–5711.
- (14) Harrison, R. S., Shepherd, N. E., Hoang, H. N., Ruiz-Gomez, G., Hill, T. A., Driver, R. W., Desai, V. S., Young, P. R., Abbenante, G., and Fairlie, D. P. (2010) Downsizing human, bacterial, and viral proteins to short water-stable α helices that maintain biological potency. *Proc. Natl. Acad. Sci. U. S. A.* 107, 11686–11691.
- (15) Hoang, H. N., Driver, R. W., Beyer, R. L., Hill, T. A., de Araujo, A. D., Plisson, F., Harrison, R. S., Goedecke, L., Shepherd, N. E., and Fairlie, D. P. (2016) Helix Nucleation by the Smallest Known α -Helix in Water. *Angew. Chem., Int. Ed.* 55, 8275–8279.
- (16) Hoang, H. N., Song, K., Hill, T. A., Derksen, D. R., Edmonds, D. J., Kok, W. M., Limberakis, C., Liras, S., Loria, P. M., Mascitti, V., Mathiowetz, A. M., Mitchell, J. M., Piotrowski, D. W., Price, D. A., Stanton, R. V., Suen, J. Y., Withka, J. M., Griffith, D. A., and Fairlie, D. P. (2015) Short hydrophobic peptides with cyclic constraints are potent glucagon-like peptide-1 receptor (GLP-1R) agonists. *J. Med. Chem.* 58, 4080–4085.
- (17) Seebach, D., and Fadel, A. (1985) N,O-Acetals from Pivalaldehyde and Amino Acids for the α -Alkylation with Self-Reproduction of the Center of Chirality. Enolates of 3-Benzoyl-2-(tert-butyl)-1,3-oxazolidin-5-ones. *Helv. Chim. Acta* 68, 1243–1250.
- (18) Mason, J. M., Schmitz, M. A., Muller, K. M., and Arndt, K. M. (2006) Semirational design of Jun-Fos coiled coils with increased affinity: Universal implications for leucine zipper prediction and design. *Proc. Natl. Acad. Sci. U. S. A.* 103, 8989–8994.
- (19) Hagemann, U. B., Mason, J. M., Müller, K. M., and Arndt, K. M. (2008) Selectional and Mutational Scope of Peptides Sequestering the Jun-Fos Coiled-Coil Domain. *J. Mol. Biol.* 381, 73–88.
- (20) Crooks, R. O., Baxter, D., Panek, A. S., Lubben, A. T., and Mason, J. M. (2016) Deriving Heterospecific Self-Assembling Protein-Protein Interactions Using a Computational Interactome Screen. *J. Mol. Biol.* 428, 385–398.
- (21) Crooks, R. O., Lathbridge, A., Panek, A. S., and Mason, J. M. (2017) Computational Prediction and Design for Creating Iteratively Larger Heterospecific Coiled Coil Sets. *Biochemistry* 56, 1573–1584.
- (22) Lathbridge, A., and Mason, J. M. (2018) Computational Competitive and Negative Design to Derive a Specific c Jun Antagonist. *Biochemistry* 57, 6108–6118.
- (23) Hill, T. A., Shepherd, N. E., Diness, F., and Fairlie, D. P. (2014) Constraining cyclic peptides to mimic protein structure motifs. *Angew. Chem., Int. Ed.* 53, 13020–13041.
- (24) Mason, J. M., Müller, K. M., and Arndt, K. M. (2007) Positive aspects of negative design: Simultaneous selection of specificity and interaction stability. *Biochemistry* 46, 4804–4814.
- (25) Mason, J. M., Müller, K. M., and Arndt, K. M. (2006) Considerations in the design and optimization of coiled coil structures. *Protein Engineering Protocols* 352, 35–70.
- (26) Mason, J. M., and Arndt, K. M. (2004) Coiled coil domains: Stability, specificity, and biological implications. *ChemBioChem* 5, 170–176.
- (27) Fletcher, J. M., Bartlett, G. J., Boyle, A. L., Danon, J. J., Rush, L. E., Lupas, A. N., and Woolfson, D. N. (2017) N@ a and N@ d:

Oligomer and Partner Specification by Asparagine in Coiled-Coil Interfaces. *ACS Chem. Biol.* 12, 528–538.

(28) Mason, J. M., Hagemann, U. B., and Arndt, K. M. (2007) Improved stability of the Jun-Fos activator protein-1 coiled coil motif: A stopped-flow circular dichroism kinetic analysis. *J. Biol. Chem.* 282, 23015–23024.

(29) Baxter, D., Perry, S. R., Hill, T. A., Kok, W. M., Zaccari, N. R., Brady, R. L., Fairlie, D. P., and Mason, J. M. (2017) Downsizing Proto-oncogene cFos to Short Helix-Constrained Peptides That Bind Jun. *ACS Chem. Biol.* 12, 2051–2061.

(30) Rao, T., Ruiz-Gómez, G., Hill, T. A., Hoang, H. N., Fairlie, D. P., and Mason, J. M. (2013) Truncated and Helix-Constrained Peptides with High Affinity and Specificity for the cFos Coiled-Coil of AP-1. *PLoS One* 8, e59415.

(31) Ganesan, S. J., and Matysiak, S. (2014) Role of Backbone Dipole Interactions in the Formation of Secondary and Super-secondary Structures of Proteins. *J. Chem. Theory Comput.* 10, 2569.

(32) Aurora, R., and Rosee, G. D. (1998) Helix capping. *Protein Sci.* 7, 21–38.

(33) Dragan, A. I., and Privalov, P. L. (2002) Unfolding of a Leucine zipper is not a Simple Two-state Transition. *J. Mol. Biol.* 321, 891–908.

(34) Worrall, J. A. R., and Mason, J. M. (2011) Thermodynamic analysis of Jun-Fos coiled coil peptide antagonists. *FEBS J.* 278, 663–672.

(35) Crick, F. H. C. (1953) The packing of α -helices: simple coiled-coils. *Acta Crystallogr.* 6, 689–697.

(36) Pauling, L., Corey, R. B., and Branson, H. R. (1951) The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. U. S. A.* 37, 205–11.

(37) Fields, G. B., and Noble, R. L. (1990) Solid phase peptide synthesis utilizing 9-fluorenylmethoxycarbonyl amino acids. *Int. J. Pept. Protein Res.* 35, 161–214.

Supporting Information

Table S1. Electrospray Mass Spectroscopy Data of Linear and Lactamised Peptides

Peptide	Sequence	Expected Linear Mass	Observed Linear Mass	Expected Cyclic Mass	Observed Cyclic Mass	Mass Change Between Linear and Cyclic
C1	YEIKAL ED	1021.50	1021.53	1003.50	1003.52	-18.01
C2	YKIKAL ED	1020.55	1020.58	1002.55	1002.57	-18.01
C3	YELKAL ED	1021.50	1021.53	1003.50	1003.53	-18.00
C4	YKLKAL ED	1021.50	1020.58	1003.50	1002.57	-18.01
C5	YEVKAL ED	1007.48	1007.51	989.48	989.51	-18.00
C6	YKVKAL ED	1006.53	1006.56	988.53	988.56	-18.00
C7	YEIKALK D	1020.55	1020.57	1002.55	1002.57	-18.00
C8	YKIKALK D	1019.60	1019.62	1001.60	1001.62	-18.00
C9	YELKALK D	1020.55	1020.58	1002.55	1002.57	-18.01
C10	YKLKALK D	1019.60	1019.62	1001.60	1001.62	-18.01
C11	YEVKALK D	1006.53	1006.57	988.53	988.55	-18.01
C12	YKVKALK D	1005.59	1005.61	987.59	987.61	-18.01
C13	YENKAL ED	1022.46	1022.48	1004.46	1004.48	-18.00
C14	YKNKAL ED	1021.51	1021.53	1003.51	1003.52	-18.01
C15	YENKALK D	1021.51	1021.53	1003.51	1003.53	-18.01
C16	YKNKALK D	1020.56	1020.59	1002.56	1002.58	-18.01

Mass spectrometry data for cassettes 1-16 showing the predicted mass, observed mass, and the difference between the linear and lactamised form of the cassettes.

Table S2. Helical and Thermal Stability of Linear and Lactamised Peptides

Peptide	fH (%)	222 nm/208 nm	T _m (°C)
cJun	29.1	0.73	25.7
1-3-16-7 Linear	65.3	0.89	29.13
1-3-16-7 Lactamised	71.4	1.01	67
6-2-13-5 Linear	18.2	0.44	N/A
6-2-13-5 Lactamised	20.4	0.69	N/A
1-3-16-7 Linear : cJun	55.2	0.87	34
1-3-16-7 Lactamised : cJun	52.0	0.90	57.2
6-2-13-5 Linear : cJun	24.9	0.60	18.9
6-2-13-5 Lactamised : cJun	25.0	0.74	29.2

A comparison of the full length peptide complexes – with their fractional helicity, 222 nm/208 nm ratio, and transition midpoint (T_m) values.

Sequence Specificity - As part of developing the modular design aspect, we attempted to gain further understanding of the ability for cassettes to tolerate lactamisation and how our computational methods could gain from this. As shown in Figure S5, we were unable to find a clear relationship between the cassettes computationally predicted to be highly stable homomerically and the helicity values found experimentally. An example of this is cassette 16, which is predicted to have low stability but has high levels of lactamised fractional helicity. Contrasting this is cassette 4, which has a high level of predicted stability but a low level of measured helicity.

As all of the cassettes exhibited an increase in helicity when lactamised, this poses an interesting question regarding the ability of certain residues to tolerate lactamisation over others. As shown in Figure 3b, there is a wide variety in the increase of helicity due to lactamisation. Though the possible permutations were limited, there is still very little that we can draw from in terms of electrostatic and core configuration to optimise the stability of the cassettes.

This presents an opportunity for the optimisation of the computational techniques employed. As previously shown,¹ bCIPA is capable of being trained on a specific subset of coiled coil forming sequences. With further experimental exploration of the cassette library, it would be possible to extract the predictors and develop a more specific prediction algorithm for this use.

Terminal Lactamisation - One of the main questions when considering multiple lactamisations within a single peptide is that of tolerance. Although it is possible that the stability increase observed in the lactamised 29-mers is due to either the N-terminal or C-terminal cassette over the other, this is something that we are able to mitigate. As shown in Figures S1 and S2, all of the peptides exhibited a shift in mass of 36 Da, indicating that both lactam bridges successfully formed between Asp and Lys, resulting in the loss of H₂O.

Although peptides truncated to below 29 residues have been explored previously²⁻⁷ with a focus on maintaining stability, it was decided against this. In keeping the modular cassette premise, a 3 cassette repeat (22-mer) would have resulted in a linear cassette flanked by two lactamised heptads. As we were exploring the effects of terminal lactamisation, it was decided that a small region of linear residues would not show us enough about the effects of the experimental technique. For this reason, a full lactamisation of the peptide was not attempted, as multiple inter-helical lactamisation events would be extremely difficult to separate from one another (and attempts to lactamise during the synthesis proved unsuccessful)

Computational Processing

All of the software was run on Python on a 64-bit x64-based processor Windows machine with 12 GB of RAM described previously¹.

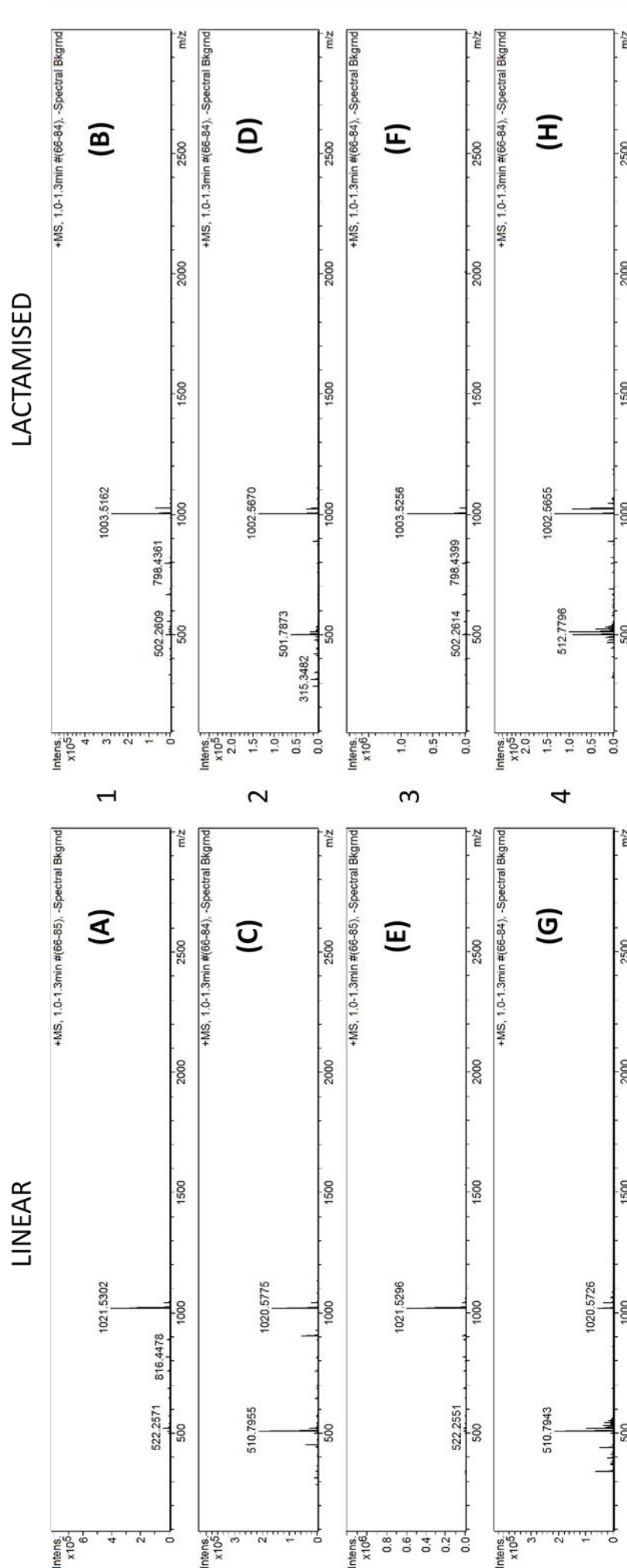


Figure S1. Electrospray Mass Spectrometry data from the purified cassettes 1-4 demonstrating a loss of 18 Da between the linear and successfully lactamised forms of cassette 1 (A/B), cassette 2 (C/D), cassette 3 (E/F), and cassette 4 (G/H)

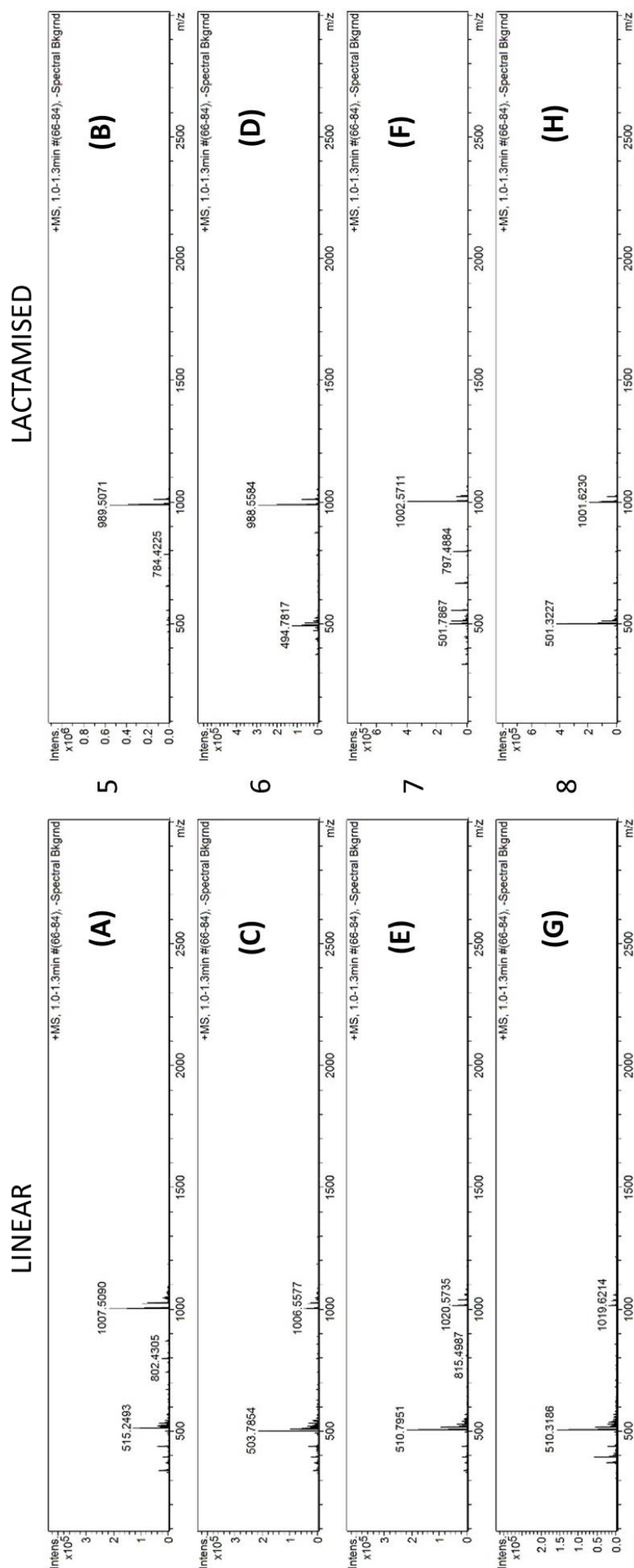


Figure S2. Electrospray Mass Spectrometry data from the purified cassettes 5-8 demonstrating a loss of 18 Da between the linear and successfully lactamised forms of cassette 5 (A/B), cassette 6 (C/D), cassette 7 (E/F), and cassette 8 (G/H)

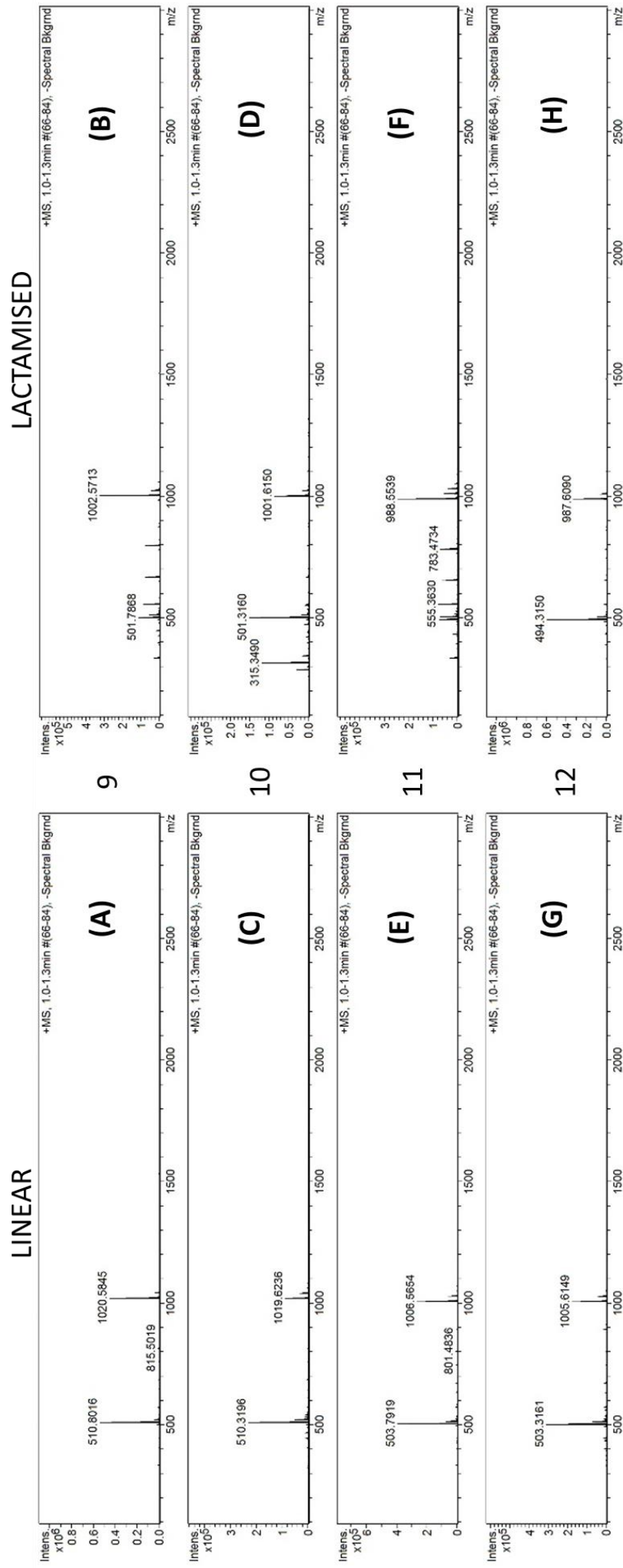


Figure S3. Electro spray Mass Spectrometry data from the purified cassettes 9-12 demonstrating a loss of 18 Da between the linear and successfully lactamised forms of cassette 9 (A/B), cassette 10 (C/D), cassette 11 (E/F), and cassette 12 (G/H)

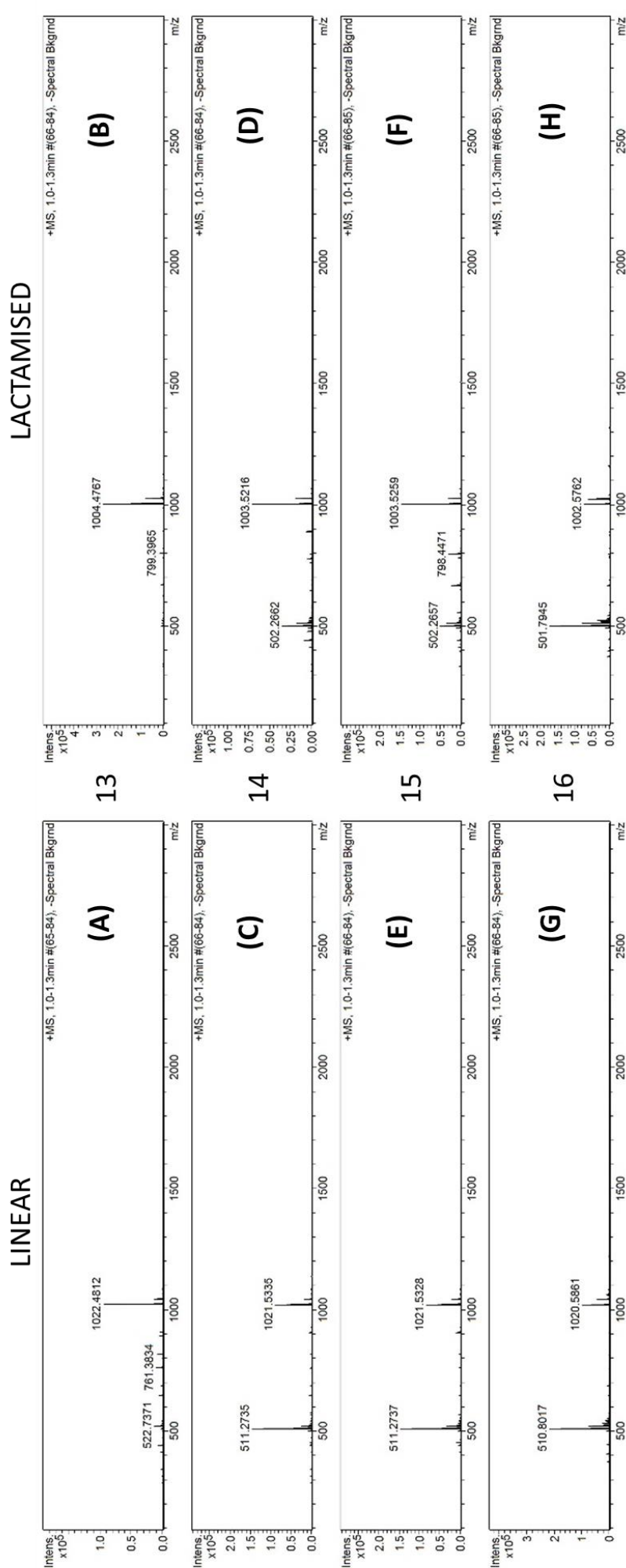


Figure S4. Electrospray Mass Spectrometry data from the purified cassettes 13-16 demonstrating a loss of 18 Da between the linear and successfully lactamised forms of cassette 13 (A/B), cassette 14 (C/D), cassette 15 (E/F), and cassette 16 (G/H)

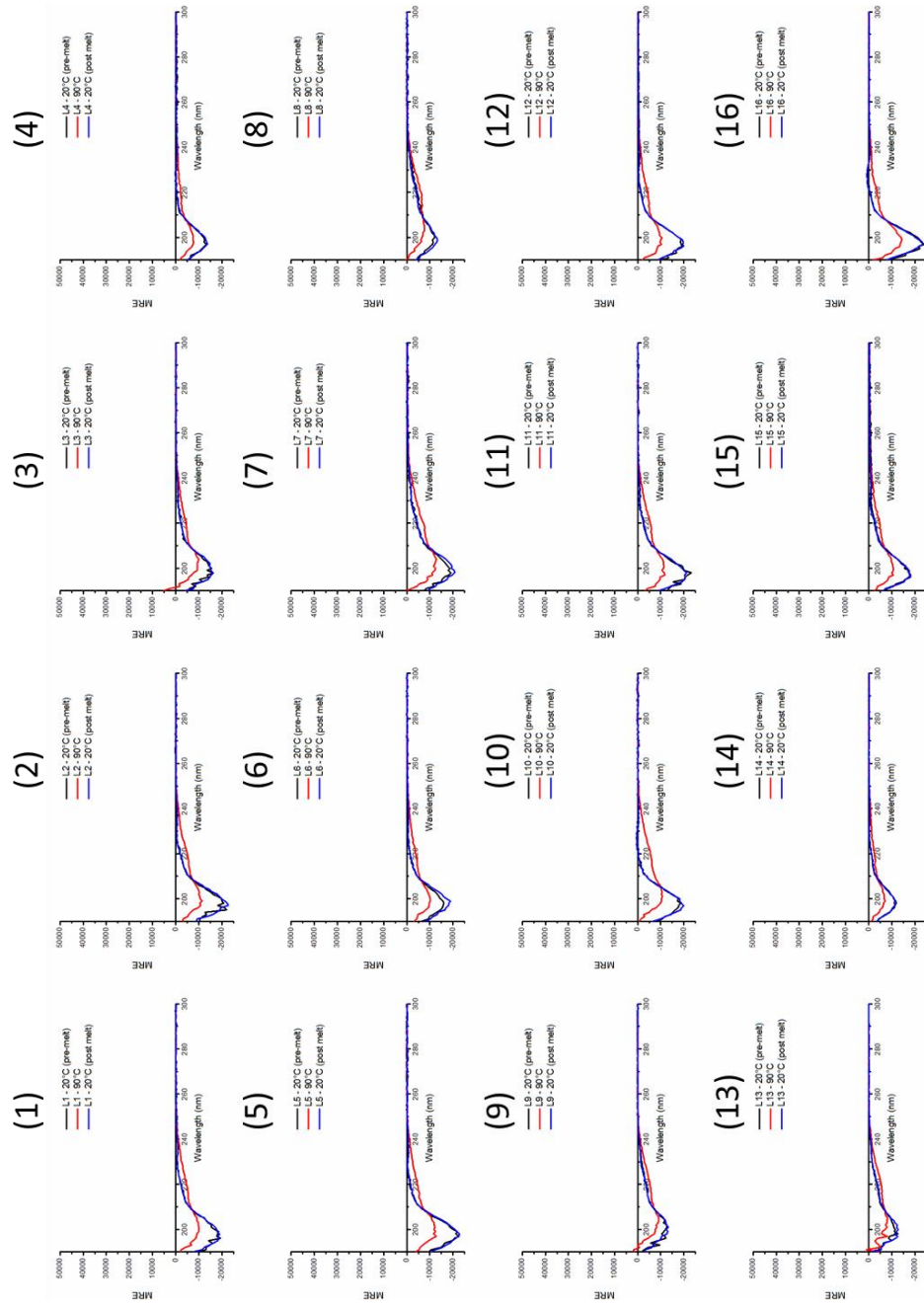


Figure S5. Thermal spectra for cassettes 1-16 in linear form. Spectra were initially measured at 20 °C (black) and in 10°C increments to 90°C (red) and then post melt at 20°C (blue) at a total peptide concentration of 150 μ M and presented as mean residue ellipticity (MRE). All experiments were performed in 10 mM potassium phosphate and 100 mM potassium fluoride (pH 7). For all cases, the linear forms were characterised as having a random coil profile at 20°C, both prior to thermal denaturation and as post-melt samples, with disorder persisting at 90°C.

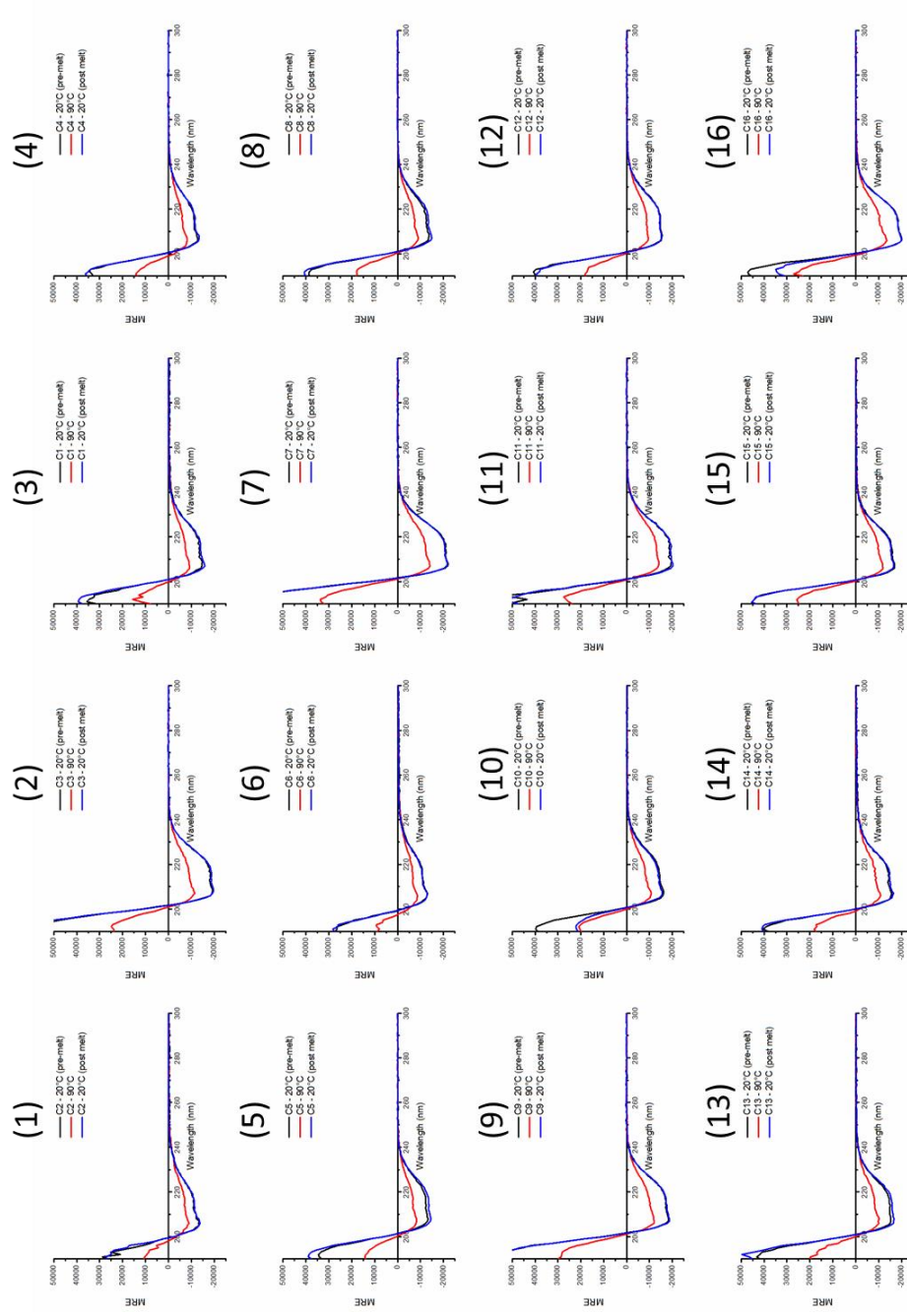


Figure S6. Thermal spectra for cassettes 1-16 in lactamised form. Spectra were initially measured at 20 °C (black) and in 10°C increments to 90°C (red) and then post melt at 20°C (blue) at a total peptide concentration of 150 µM and presented as mean residue ellipticity (MRE). All experiments were performed in 10 mM potassium phosphate and 100 mM potassium fluoride (pH 7). In all cases, lactamised cassettes presented as α -helical at 20°C and returned to within 5% of the same helical signature as post-melt samples, with an average helicity of 48%. Consistent with lactamisation as a potent helix-inducer, at 90°C helicity was retained with an average fH of 38%.

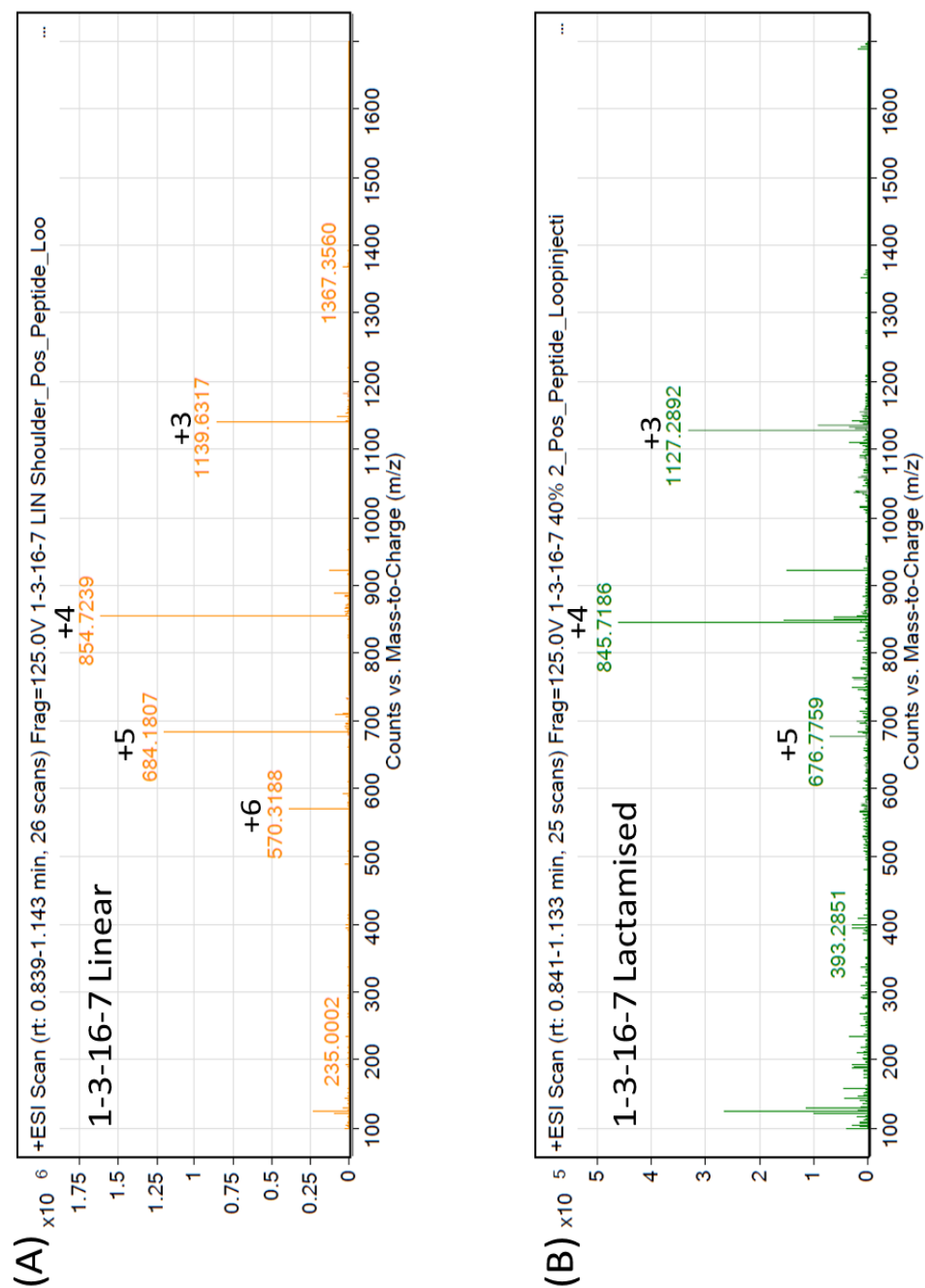


Figure S7. Electrospray Mass Spectrometry data from the purified full-length 1-3-16-7 demonstrates the peptide to have the expected mass of 3415.91 Da as a linear sequence (A) and the expected mass of 3379.91 Da when lactamised (B). The loss of 36 Da demonstrates that there has been successful lactamisation at both termini consistent with the loss of two water molecules.

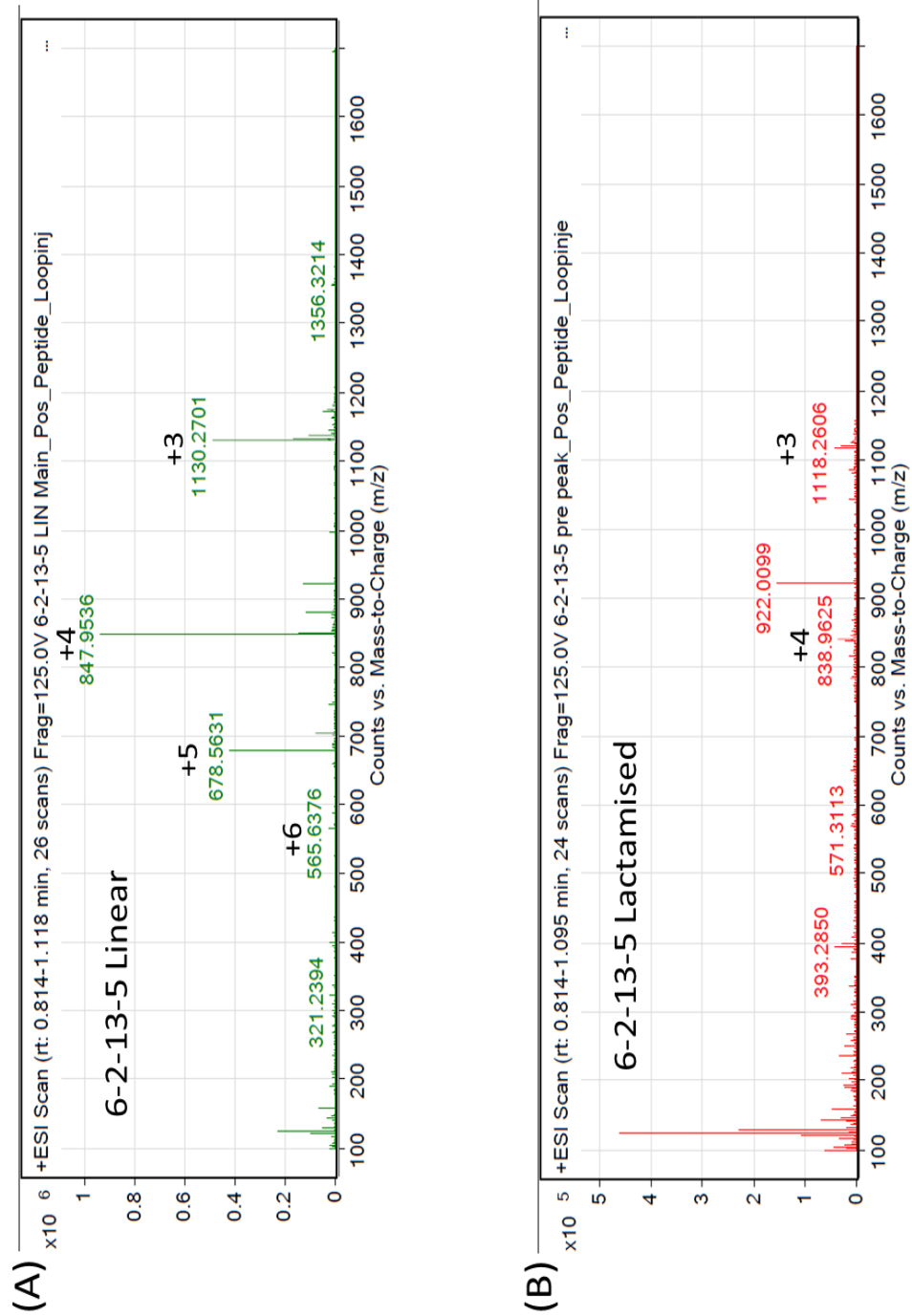


Figure S8. Electrospray Mass Spectrometry data from the purified full-length 6-2-13-5 demonstrates the peptide to have the expected mass of 3388.79 Da as a linear sequence (A) and the expected mass of 3352.79 Da when lactamised (B). The loss of 36 Da demonstrates that there has been successful lactamisation at both termini consistent with the loss of two water molecules.

References

- (1) Lathbridge, A., and Mason, J. M. (2018) Computational Competitive and Negative Design to Derive a Specific cJun Antagonist. *Biochemistry* 57, 6108–6118.
- (2) Rao, T., Ruiz-Gómez, G., Hill, T. A., Hoang, H. N., Fairlie, D. P., and Mason, J. M. (2013) Truncated and Helix-Constrained Peptides with High Affinity and Specificity for the cFos Coiled-Coil of AP-1. *PLoS One* 8, e59415.
- (3) Baxter, D., Perry, S. R., Hill, T. A., Kok, W. M., Zaccari, N. R., Brady, R. L., Fairlie, D. P., and Mason, J. M. (2017) Downsizing Proto-oncogene cFos to Short Helix-Constrained Peptides That Bind Jun. *ACS Chem. Biol.* 12, 2051–2061.
- (4) Hoang, H. N., Driver, R. W., Beyer, R. L., Hill, T. A., D. de Araujo, A., Plisson, F., Harrison, R. S., Goedecke, L., Shepherd, N. E., and Fairlie, D. P. (2016) Helix Nucleation by the Smallest Known α -Helix in Water. *Angew. Chemie - Int. Ed.* 55, 8275–8279.
- (5) De Araujo, A. D., Hoang, H. N., Kok, W. M., Diness, F., Gupta, P., Hill, T. A., Driver, R. W., Price, D. A., Liras, S., and Fairlie, D. P. (2014) Comparative α -helicity of cyclic pentapeptides in water. *Angew. Chemie - Int. Ed.* 53, 6965–6969.
- (6) Shepherd, N. E., Abbenante, G., and Fairlie, D. P. (2004) Consecutive cyclic pentapeptide modules form short α -helices that are very stable to water and denaturants. *Angew. Chemie - Int. Ed.* 43, 2687–2690.
- (7) Mazzier, D., Peggion, C., Toniolo, C., and Moretto, A. (2014) Enhancement of the helical content and stability induced in a linear oligopeptide by an i, i+4 intramolecularly double stapled, overlapping, bicyclic [31, 22, 5]-(E)ene motif. *Biopolym. - Pept. Sci. Sect.* 102, 115–123.

CHAPTER 7 - CONCLUSION

7.1: *In Silico* Screening of Coiled Coils

The overarching goal of this thesis was to understand and optimise the ability to address the conflicting design requirements posed by the AP-1 system: engineering the stabilisation of the single desired interaction with concurrent destabilisation of the multiple off-target interactions.

The choice to use cFos as a template for antagonists has built on the previous work explored in this group. It represents a far better antagonist – target interaction, given the ability of cJun to dimerise AP-1. As highlighted, Fos-based antagonists also have the ability to maximise the effect of the Δ value used during the *in silico* screening – with both FosU_{isCAN} and FosU_{PCA} presenting with stable desired complexes and unstable off-target complexes. This highlights the effect of basing an antagonist on an unstable peptide; contrasting with previously characterised peptides JunW and JunW_{CANDI} – both of which display stable homodimeric complexes. Due to the residues at the interfacial positions and those found within the hydrophobic interface, the ability of the Fos-based peptides to form a stable off-target complex is limited (due to their optimisation for cJun).

As it is understood, the evolution of *in silico* design and screening is influenced by the experimental data available. As discussed in Chapter 4 and Chapter A1, the flexibility of the base bCIPA algorithm allows for different predictors to be utilised within the *in silico* system when applied to experimental data from less diverse systems. The development of isPCA and isCAN has explored many novel areas in the design of computational frameworks for CC prediction. The use of the Δ parameter can be seen as similar to the stability/specificity trade-off found in CLASSY (whose features are discussed in Chapter 1), however the use of BOW in an analogous method would be extremely cumbersome (Fong et al. 2004; Grigoryan et al. 2009b). This is due to the inherent difficulty of using categorical data in complex combinations with desired specificity parameterised in Δ . In comparison to CLASSY, the advancement of these bCIPA-based frameworks represents combined novelty and simplicity. These tools have the capacity to screen large libraries without requiring simultaneous *in vitro* validation, meaning that the factors resulting in the success and failure of certain library members can be elucidated. As highlighted in Chapter 6, the expansion of isCAN into predicting two separate libraries to be screened against one another (as explained in Figure 6.11)

This highlights the ability for this framework to be further tailored to specific coiled coil subsets for which further nuance is required while screening. Although this has been shown

to provide a greater level of accuracy, it does not (at this stage) elucidate any more about the rules governing the interaction of coiled coils – merely expanding properties already incorporated into the other predictors.

The use of *in cellulo* PCA screening resulted in the selection of a peptide sequence not highly ranked within the *in silico* screen. In combination with the rest of this thesis, this has raised a separate and more fundamental question about the use of pairwise prediction. With bCIPA scoring each residue independent of others intra-helically (with inter-helical scoring at the interfaces), this takes advantage of the coupling data used to create it. However, this fails to take into account the local effects of the residues at each position. Work exploring this (Potapov et al. 2015) has highlighted the limitation of relying solely on coupling data. Although the findings promote the use of a model incorporating further parameters to encompass these effects, it found that the weighting of this model correlated with the binding data used to develop bCIPA (Krylov et al. 1998a) – not completely disregarding the ability of bCIPA as a data-driven prediction model.

If for some cases, the predicted T_m from a single bCIPA pairwise screen does not accurately map to the experimental value, is there an inherent flaw in designing complex *in silico* methodologies based on the combination of a multitude of these predictions? As this thesis has argued, the ability to impose “artificial” limits through the Δ parameter mitigates this somewhat – by replacing the subtle differences of desired/undesired predictions with predictably larger variations. This does not indicate that this approach is ill-suited to CC screening – it exists as a necessary safeguard for the additional layers of complexity employed.

This supports the view that – given an established template – it is possible to design a sequence *in silico* with comparable ability (in terms of affinity) to that of one derived from *in cellulo* screening techniques. While it does not represent a way by which *in silico* can be used solely for peptide design, it does present a step forward in the understanding of specificity as it relates to the conflicting design requirements.

7.2: Truncation and Elongation

The exploration of modular design in Chapter 5 highlighted the fallacy of relying solely on *in silico* prediction methods for short sequences while also highlighting the ability of the truncated but stabilised peptides to have high levels of engineered stability from data combining individual sequences. This allows for future exploration into the link between the helicity of individual heptads and that of full-length sequences. A benefit of utilising a smaller library is the limited diversity in randomised options and this lends itself to further research into sequence design as it relates to engineering tolerance of the lactam bridge.

As shown by the normalisation employed within the study, prediction on truncated seven residue cassettes results not in standard T_m values but basic quantitative data. As sequences of this length are not able to natively form α -helices (and coiled coils, by extension), this is not a flaw in the methodology and can be attributed to discrepancy between the length of the heptad cassettes and that of the peptide sequences used to train bCIPA. Although this is the case, the prediction calculations themselves are not altered fundamentally – it is still the summation of helical propensity, core, and electrostatic scoring. However, values received from short sequences like this are nonetheless very powerful in providing rank order of predicted affinities.

While our understanding of the effects of lactamisation is wide-ranging, current knowledge on the environment required for tolerating this is limited. As an application of the current work, it would be possible to create a limited comparative database of ‘cassettes’, with relative scoring. This would allow for further development of this approach as an example of data-driven design, with the experimental data refining the ability to design heptads of highest stability, highest tolerance of lactamisation, and the optimal combinations of heptads (linear and lactamised) to design a full-length sequence. Although cJun was utilised as an example heteromeric interaction, the aforementioned data-driven design approach would allow for design of potentially novel antagonists. As the library was not optimised for targeting cJun but still displayed heteromeric stability with a full-length sequence, it could be that is that residue selection (even with an expansive heptad library) using this modular approach would differ from previously designed sequences (combining residue-residue scoring with the optimisation of lactamisation). This means that, while tentative, the work described represents a step forward in the application of bCIPA to modular *de novo* peptide design.

The elongation of the sequences involved in Chapter 4 and 5 from 37-mers to 39-mers resulted in additional electrostatic interactions due to extra **g** and **e** position residues. From the observations of reciprocal and unbalanced extension (the latter present with non-extended JunW_{CANDI}), the respective addition of 4 and 2 extra interactions resulted in a greater difference between the T_m values of the desired and undesired complexes. This can be ascribed to the role of the electrostatic interactions, with elongation increasing the specificity of the CCs. The additional stability conferred to homomeric cJun was similarly driven by this. Further work in this area would involve the creation of dual cFos and cJun targeting libraries which consider each other as competitors to develop a pair of AP-1 antagonists which are co-compatible.

When using isPCA/isCAN to develop the antagonist peptides, the base template used was that of cFos – with randomisation at **a**, **e**, and **g** positions. Discounting the lack of modification of the **d** position from Leu, the heptad backbone (**b**, **c**, and **f**) was not included in the semi-rational design of the sequences – choosing to keep them as the residues found

in cFos. Our limited understanding of these positions is reflected within bCIPA – in which the residues at these **b**, **c**, and **f** positions contribute solely to the helical propensity calculation. As it is thought that these positions do not play a role in the inter-helical interaction, randomisation at these positions was not considered a priority due to the lack of data within the underlying bCIPA algorithm.

7.3: Summary Conclusions

The summary of the overall thesis conclusions are as follows:

- Creating *in silico* homologues of PCA and CANDI as isPCA and isCAN allows for greater screening potential, allowing the user to create much larger *in silico* libraries than can be paired down into high predicted quality libraries that are accessible to *in cellulo* screening. The *in silico* library size is only limited by computational performance.
- The use of the Δ parameter within these *in silico* techniques allows an increased level of stringency, resulting in the selection of highly specific peptides.
- The extension of peptide sequences to incorporate an additional 4 electrostatic interactions increases both stability and the specificity of the peptide complex – allowing for flexibility to form repulsive interactions with off-targets and beneficial interactions for desired complexes.
- The coupling of *in silico* and *in cellulo* screening techniques results in sequences that have similar stability to purely *in cellulo* screened peptides but with lower levels of stability with off-target complexes – with *in silico* methods acting as a pre-screen for the limited size PCA.
- Using *in vitro* measurement of global secondary structure, it is possible to create a library of short 7-mer heptad linear/lactamised cassettes which can be combined to form four heptad peptides whose stability reflects the helicity of the individual cassettes.
- Lactamisation of truncated peptides increased the stability relative to the linear forms, and resulted in peptides that were able to successfully bind to cJun with varying levels of stability reflecting the helicity of their individual cassettes.
- Using a similar but numerous data-set, it is possible to introduce additional weighted predictors into bCIPA to develop a prediction model with increased accuracy (qCIPA). The main limitations of this approach is the availability of experimental data and amount of diversity within the set.
- The use of *in silico* techniques to design and screen sequences *de novo* is not yet possible but this is predicted to become achievable with an increase of experimental data.

Thesis References

- Acharya, A., Rishi, V. and Vinson, C., 2006. Stability of 100 homo and heterotypic coiled-coil a-a' pairs for ten amino acids (A, L, I, V, N, K, S, T, E, and R). *Biochemistry*, 45(38), pp.11324–11332. Available from: <https://doi.org/10.1021/bi060822u> [Accessed 28 September 2016].
- Aihara, K., Inokuma, T., Komiya, C., Shigenaga, A. and Otaka, A., 2015. Synthesis of lactam-bridged cyclic peptides using sequential olefin metathesis and diimide reduction reactions. *Tetrahedron*, 71(24), pp.4183–4191. Available from: <https://doi.org/10.1016/j.tet.2015.04.093>.
- Alber, T., 1992. Structure of the leucine zipper. *Current Opinion in Genetics and Development*, 2(2), pp.205–210. Available from: [https://doi.org/10.1016/S0959-437X\(05\)80275-8](https://doi.org/10.1016/S0959-437X(05)80275-8).
- Alfonso-Jaume, M.A., Bergman, M.R., Mahimkar, R., Cheng, S., Jin, Z.Q., Karliner, J.S. and Lovett, D.H., 2006. Cardiac ischemia-reperfusion injury induces matrix metalloproteinase-2 expression through the AP-1 components FosB and JunB. *American Journal of Physiology - Heart and Circulatory Physiology*, 291(4). Available from: <https://doi.org/10.1152/ajpheart.00026.2006>.
- Angel, P. and Karin, M., 1991. The role of Jun, Fos and the AP-1 complex in cell-proliferation and transformation. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1072(2–3), pp.129–157. Available from: [https://doi.org/10.1016/0304-419X\(91\)90011-9](https://doi.org/10.1016/0304-419X(91)90011-9) [Accessed 26 September 2016].
- Apostolou, P., Toloudi, M., Ioannou, E., Chatziioannou, M., Kourtidou, E., Vlachou, I. and Papasotiriou, I., 2013. AP-1 Gene Expression Levels May Be Correlated with Changes in Gene Expression of Some Stemness Factors in Colon Carcinomas. *Journal of signal transduction*, 2013, p.497383. Available from: <https://doi.org/10.1155/2013/497383> [Accessed 8 January 2018].
- Babu, R.L., Naveen Kumar, M., Patil, R.H., Devaraju, K.S., Ramesh, G.T. and Sharma, S.C., 2013. Effect of estrogen and tamoxifen on the expression pattern of AP-1 factors in MCF-7 cells: Role of c-Jun, c-Fos, and Fra-1 in cell cycle regulation. *Molecular and Cellular Biochemistry*, 380(1–2), pp.143–151. Available from: <https://doi.org/10.1007/s11010-013-1667-x>.
- Bakin, A. V. and Curran, T., 1999. Role of DNA 5-methylcytosine transferase in cell transformation by fos. *Science*, 283(5400), pp.387–390. Available from: <https://doi.org/10.1126/science.283.5400.387>.

- Ball, D.P., Lewis, A.M., Williams, D., Resetca, D., Wilson, D.J. and Gunning, P.T., 2016. Signal transducer and activator of transcription 3 (STAT3) inhibitor, S3I-201, acts as a potent and non-selective alkylating agent. *Oncotarget*, 7(15). Available from: <https://doi.org/10.18632/oncotarget.7838>.
- Basbous, J., Jariel-Encontre, I., Gomard, T., Bossis, G. and Piechaczyk, M., 2008. Ubiquitin-independent- versus ubiquitin-dependent proteasomal degradation of the c-Fos and Fra-1 transcription factors: Is there a unique answer? *Biochimie*, 90(2), pp.296–305. Available from: <https://doi.org/10.1016/J.BIOCHI.2007.07.016> [Accessed 13 May 2019].
- Batley, J., Moulding, C., Taub, R., Murphy, W., Stewart, T., Potter, H., Lenoir, G. and Leder, P., 1983. The human c-myc oncogene: Structural consequences of translocation into the igh locus in Burkitt lymphoma. *Cell*, 34(3), pp.779–787. Available from: [https://doi.org/10.1016/0092-8674\(83\)90534-2](https://doi.org/10.1016/0092-8674(83)90534-2).
- Baxter, D., Perry, S.R., Hill, T.A., Kok, W.M., Zaccai, N.R., Brady, R.L., Fairlie, D.P. and Mason, J.M., 2017. Downsizing Proto-oncogene cFos to Short Helix-Constrained Peptides That Bind Jun. *ACS Chemical Biology*, 12(8), pp.2051–2061. Available from: <https://doi.org/10.1021/acschembio.7b00303>.
- Baxter, D., Ullman, C.G., Frigotto, L. and Mason, J.M., 2017. Exploiting Overlapping Advantages of In Vitro and In Cellulo Selection Systems to Isolate a Novel High-Affinity cJun Antagonist. *ACS Chemical Biology*, 12(10), pp.2579–2588. Available from: <https://doi.org/10.1021/acschembio.7b00693> [Accessed 10 January 2018].
- Beesley, J.L. and Woolfson, D.N., 2019. The de novo design of α -helical peptides for supramolecular self-assembly. *Current Opinion in Biotechnology*, 58, pp.175–182. Available from: <https://doi.org/10.1016/j.copbio.2019.03.017>.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G. and Bhat, T. N. & Weissig, H., 2000. The Protein Data Bank. *Nucleic Acids Research*, 28(1), pp.235–242. Available from: <https://doi.org/10.1093/nar/28.1.235> [Accessed 20 August 2019].
- Bromley, E.H.C., Sessions, R.B., Thomson, A.R. and Woolfson, D.N., 2009. Designed α -helical tectons for constructing multicomponent synthetic biological systems. *Journal of the American Chemical Society*, 131(3), pp.928–930. Available from: <https://doi.org/10.1021/ja804231a>.
- Bruzzoni-Giovanelli, H., Alezra, V., Wolff, N., Dong, C.Z., Tuffery, P. and Rebollo, A., 2018.

Interfering peptides targeting protein–protein interactions: the next generation of drugs? *Drug Discovery Today*, 23(2), pp.272–285. Available from: <https://doi.org/10.1016/j.drudis.2017.10.016>.

Burges, C.J.C., 1998. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 2(2), pp.121–167. Available from: <https://doi.org/10.1023/A:1009715923555> [Accessed 21 August 2019].

Caput, D., Beutler, B., Hartog, K., Thayer, R., Brown-Shimer, S. and Cerami, A., 1986. Identification of a common nucleotide sequence in the 3'-untranslated region of mRNA molecules specifying inflammatory mediators. *Proceedings of the National Academy of Sciences of the United States of America*, 83(6), pp.1670–4. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/2419912> <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC323145>.

Caputto, B.L., Cardozo Gizzi, A.M. and Gil, G.A., 2014. c-Fos: An AP-1 transcription factor with an additional cytoplasmic, non-genomic lipid synthesis activation capacity. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, 1841(9), pp.1241–1246. Available from: <https://doi.org/10.1016/J.BBALIP.2014.05.007> [Accessed 24 April 2019].

Chakraborti, S., Mandal, M., Das, S., Mandal, A. and Chakraborti, T., 2003. Regulation of matrix metalloproteinases. An overview. *Molecular and Cellular Biochemistry*, 253(1–2), pp.269–285. Available from: <https://doi.org/10.1023/A:1026028303196>.

Chapman, R.N., Dimartino, G. and Arora, P.S., 2004. A highly stable short α -helix constrained by a main-chain hydrogen-bond surrogate. *Journal of the American Chemical Society*, 126(39), pp.12252–12253. Available from: <https://doi.org/10.1021/ja0466659>.

Che, Y., Brooks, B.R. and Marshall, G.R., 2006. Development of small molecules designed to modulate protein-protein interactions. *Journal of Computer-Aided Molecular Design*, 20(2), pp.109–130. Available from: <https://doi.org/10.1007/s10822-006-9040-8>.

Chen, G. and Goeddel, D. V., 2002. TNF-R1 signaling: A beautiful pathway. *Science*, 296(5573), pp.1634–1635. Available from: <https://doi.org/10.1126/science.1071924>.

Chen, T.S., Reinke, A.W. and Keating, A.E., 2011. Design of peptide inhibitors that bind the bZIP domain of Epstein-Barr virus protein BZLF1. *Journal of Molecular Biology*, 408(2), pp.304–320. Available from: <https://doi.org/10.1016/j.jmb.2011.02.046> [Accessed 27 June 2016].

Chen, Y.-L., Chan, S.-H., Lin, P.-Y. and Chu, P.-Y., 2017. The expression of a tumor

- suppressor gene JDP2 and its prognostic value in hepatocellular carcinoma patients ☆,☆☆. *Human Pathology*, 63, pp.212–216. Available from: <https://doi.org/10.1016/j.humpath.2017.03.003> [Accessed 25 April 2019].
- Chen, Y., Mauldin, J.P., Day, R.N. and Periasamy, A., 2007. Characterization of spectral FRET imaging microscopy for monitoring nuclear protein interactions. *Journal of microscopy*, 228(Pt 2), pp.139–52. Available from: <https://doi.org/10.1111/j.1365-2818.2007.01838.x> [Accessed 19 August 2019].
- Chinenov, Y. and Kerppola, T.K., 2001. Close encounters of many kinds: Fos-Jun interactions that mediate transcription regulatory specificity. *Oncogene*, 20(19), pp.2438–2452. Available from: <https://doi.org/10.1038/sj.onc.1204385> [Accessed 27 August 2019].
- Chothia, C., Levitt, M. and Richardson, D., 1977. Structure of proteins: packing of alpha-helices and pleated sheets. *Proceedings of the National Academy of Sciences of the United States of America*, 74(10), p.4130. Available from: <https://doi.org/10.1073/PNAS.74.10.4130> [Accessed 7 May 2019].
- Chothia, C., Levitt, M. and Richardson, D., 1981. Helix to helix packing in proteins. *Journal of Molecular Biology*, 145(1), pp.215–250. Available from: [https://doi.org/10.1016/0022-2836\(81\)90341-7](https://doi.org/10.1016/0022-2836(81)90341-7) [Accessed 7 May 2019].
- Craik, D.J., Fairlie, D.P., Liras, S. and Price, D., 2013. The Future of Peptide-based Drugs. *Chemical Biology and Drug Design*, 81(1), pp.136–147. Available from: <https://doi.org/10.1111/cbdd.12055>.
- Crick, F.H.C., 1953. *The packing of α -helices: simple coiled-coils*. International Union of Crystallography (IUCr). Available from: <https://doi.org/doi:10.1107/S0365110X53001964> [Accessed 29 April 2019].
- Crooks, R.O., Baxter, D., Panek, A.S., Lubben, A.T. and Mason, J.M., 2016. Deriving Heterospecific Self-Assembling Protein-Protein Interactions Using a Computational Interactome Screen. *Journal of Molecular Biology*, 428(2), pp.385–398. Available from: <https://doi.org/10.1016/j.jmb.2015.11.022> [Accessed 20 September 2016].
- Crooks, R.O., Lathbridge, A., Panek, A.S. and Mason, J.M., 2017. Computational Prediction and Design for Creating Iteratively Larger Heterospecific Coiled Coil Sets. *Biochemistry*, 56(11), pp.1573–1584. Available from: <https://doi.org/10.1021/acs.biochem.7b00047> [Accessed 31 January 2018].
- Crooks, R.O., Rao, T. and Mason, J.M., 2011a. Truncation, randomization, and selection: Generation of a reduced length c-jun antagonist that retains high interaction stability.

Journal of Biological Chemistry, 286(34), pp.29470–29479. Available from: <https://doi.org/10.1074/jbc.M111.221267> [Accessed 19 September 2015].

Crooks, R.O., Rao, T. and Mason, J.M., 2011b. Truncation, randomization, and selection: Generation of a reduced length c-jun antagonist that retains high interaction stability. *Journal of Biological Chemistry*, 286(34), pp.29470–29479. Available from: <https://doi.org/10.1074/jbc.M111.221267> [Accessed 19 September 2015].

Cumaraswamy, A.A., Lewis, A.M., Geletu, M., Todici, A., Diaz, D.B., Cheng, X.R., Brown, C.E., Laister, R.C., Muench, D., Kerman, K., Grimes, H.L., Minden, M.D. and Gunning, P.T., 2014. Nanomolar-potency small molecule inhibitor of STAT5 protein. *ACS Medicinal Chemistry Letters*, 5(11), pp.1202–1206. Available from: <https://doi.org/10.1021/ml500165r>.

Curran, T. and Morgan, J.I., 1987. Memories of fos. *BioEssays*, 7(6), pp.255–258. Available from: <https://doi.org/10.1002/bies.950070606>.

Dagleish, G., Veyrune, J.-L., Blanchard, J.-M. and Hesketh, J., 2001. mRNA Localization by a 145-Nucleotide Region of the c- fos 3'- Untranslated Region. *Journal of Biological Chemistry*, 276(17), pp.13593–13599. Available from: <https://doi.org/10.1074/jbc.M001141200>.

van Dam, H. and Castellazzi, M., 2001. Distinct roles of Jun : Fos and Jun : ATF dimers in oncogenesis. *Oncogene*, 20(19), pp.2453–2464. Available from: <https://doi.org/10.1038/sj.onc.1204239> [Accessed 10 May 2019].

Darlyuk-Saadon, I., Weidenfeld-Baranboim, K., Yokoyama, K.K., Hai, T. and Aronheim, A., 2012. The bZIP repressor proteins, c-Jun dimerization protein 2 and activating transcription factor 3, recruit multiple HDAC members to the ATF3 promoter. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1819(11–12), pp.1142–1153. Available from: <https://doi.org/10.1016/j.bbagr.2012.09.005> [Accessed 27 June 2016].

Darnell, J.E., 2002. Transcription factors as targets for cancer therapy. *Nature Reviews Cancer*, 2(10), pp.740–749. Available from: <https://doi.org/10.1038/nrc906> [Accessed 23 June 2017].

Dedieu, S. and Lefebvre, P., 2006. Retinoids interfere with the AP1 signalling pathway in human breast cancer cells. *Cellular Signalling*, 18(6), pp.889–898. Available from: <https://doi.org/10.1016/j.cellsig.2005.08.001>.

Deppmann, C.D., Alvania, R.S. and Taparowsky, E.J., 2006. Cross-Species Annotation of Basic Leucine Zipper Factor Interactions: Insight into the Evolution of Closed

- Interaction Networks. *Molecular Biology and Evolution*, 23(8), pp.1480–1492. Available from: <https://doi.org/10.1093/molbev/msl022> [Accessed 7 May 2019].
- Dérjard, B., Hibi, M., Wu, I.-H., Barrett, T., Su, B., Deng, T., Karin, M. and Davis, R.J., 1994. JNK1: A protein kinase stimulated by UV light and Ha-Ras that binds and phosphorylates the c-Jun activation domain. *Cell*, 76(6), pp.1025–1037. Available from: [https://doi.org/10.1016/0092-8674\(94\)90380-8](https://doi.org/10.1016/0092-8674(94)90380-8).
- Eckert, R.L., Adhikary, G., Young, C.A., Jans, R., Crish, J.F., Xu, W. and Rorke, E.A., 2013. AP1 Transcription Factors in Epidermal Differentiation and Skin Cancer. *Journal of Skin Cancer*, 2013, pp.1–9. Available from: <https://doi.org/10.1155/2013/537028> [Accessed 25 April 2019].
- Eferl, R. and Wagner, E.F., 2003. AP-1: A double-edged sword in tumorigenesis. *Nature Reviews Cancer*, 3(11), pp.859–868. Available from: <https://doi.org/10.1038/nrc1209> [Accessed 24 May 2016].
- Efimov, A. V., 1999. Complementary packing of α -helices in proteins. *FEBS Letters*, 463(1–2), pp.3–6. Available from: [https://doi.org/10.1016/S0014-5793\(99\)01507-0](https://doi.org/10.1016/S0014-5793(99)01507-0) [Accessed 7 May 2019].
- Ellert-Miklaszewska, A., Poleszak, K. and Kaminska, B., 2017. Short peptides interfering with signaling pathways as new therapeutic tools for cancer treatment. *Future Medicinal Chemistry*, 9(2), pp.199–221. Available from: <https://doi.org/10.4155/fmc-2016-0189>.
- Endo, S., Fujita, M., Yamada, S., Imadome, K., Nakayama, F., Isozaki, T., Yasuda, T., Imai, T. and Matsubara, H., 2018. Fra-1 enhances the radioresistance of colon cancer cells to X-ray or C-ion radiation. *Oncology Reports*, 39(3), pp.1112–1118. Available from: <https://doi.org/10.3892/or.2018.6223>.
- Farley, K.A., Che, Y., Navarro-Vázquez, A., Limberakis, C., Anderson, D., Yan, J., Shapiro, M., Shanmugasundaram, V. and Gil, R.R., 2019. Cyclic Peptide Design Guided by Residual Dipolar Couplings, J-Couplings, and Intramolecular Hydrogen Bond Analysis. *Journal of Organic Chemistry*, 84(8), pp.4803–4813. Available from: <https://doi.org/10.1021/acs.joc.8b02811>.
- Fields, G.B. and Noble, R.L., 1990. Solid phase peptide synthesis utilizing 9-fluorenylmethoxycarbonyl amino acids. *International journal of peptide and protein research*, 35(3), pp.161–214. Available from: <https://doi.org/10.1111/j.1399-3011.1990.tb00939.x>.
- Fletcher, J.M., Bartlett, G.J., Boyle, A.L., Danon, J.J., Rush, L.E., Lupas, A.N. and

- Woolfson, D.N., 2016. N@a and N@d: Oligomer and Partner Specification by Asparagine in Coiled-Coil Interfaces. *ACS Chemical Biology*. Available from: <https://doi.org/10.1021/acschembio.6b00935> [Accessed 9 January 2017].
- Fletcher, J.M., Bartlett, G.J., Boyle, A.L., Danon, J.J., Rush, L.E., Lupas, A.N. and Woolfson, D.N., 2017. N@ a and N@ d: Oligomer and Partner Specification by Asparagine in Coiled-Coil Interfaces. *ACS Chemical Biology*, 12(2), pp.528–538. Available from: <https://doi.org/10.1021/acschembio.6b00935> [Accessed 9 January 2017].
- Fong, J., Keating, A. and Singh, M., 2004. Predicting specificity in bZIP coiled-coil protein interactions. *Genome Biology*, 5(2), p.R11. Available from: <https://doi.org/10.1186/gb-2004-5-2-r11> [Accessed 13 January 2016].
- Garces de Los Fayos Alonso, I., Liang, H.C., Turner, S.D., Lager, S., Merkel, O. and Kenner, L., 2018. The role of activator protein-1 (AP-1) family members in CD30-positive lymphomas. *Cancers*, 10(4), pp.1–22. Available from: <https://doi.org/10.3390/cancers10040093>.
- Glover, J.N.M. and Harrison, S.C., 1995. Crystal structure of the heterodimeric bZIP transcription factor c-Fos–c-Jun bound to DNA. *Nature*, 373(6511), pp.257–261. Available from: <https://doi.org/10.1038/373257a0> [Accessed 25 April 2019].
- Graddis, T.J., Myszka, D.G. and Chaiken, I.M., 1993. Controlled Formation of Model Homo- and Heterodimer Coiled Coil Polypeptides. *Biochemistry*, 32(47), pp.12664–12671. Available from: <https://doi.org/10.1021/bi00210a015>.
- Gradišar, H., Božič, S., Doles, T., Vengust, D., Hafner-Bratkovič, I., Mertelj, A., Webb, B., Šali, A., Klavžar, S. and Jerala, R., 2013. Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments. *Nature Chemical Biology*, 9(6), pp.362–366. Available from: <https://doi.org/10.1038/nchembio.1248>.
- Grigoryan, G. and Keating, A.E., 2006. Structure-based prediction of bZIP partnering specificity. *Journal of Molecular Biology*, 355(5), pp.1125–1142. Available from: <https://doi.org/10.1016/j.jmb.2005.11.036>.
- Grigoryan, G., Reinke, A.W. and Keating, A.E., 2009a. Design of protein-interaction specificity affords selective bZIP-binding peptides. *Nature*, 458(7240), pp.859–864. Available from: <https://doi.org/10.1038/nature07885> [Accessed 27 June 2016].
- Grigoryan, G., Reinke, A.W. and Keating, A.E., 2009b. Design of protein-interaction specificity gives selective bZIP-binding peptides. *Nature*, 458(7240), pp.859–864. Available from: <https://doi.org/10.1038/nature07885> [Accessed 27 June 2016].

- Gungl, A., Biasin, V., Wilhelm, J., Olschewski, A., Kwapiszewska, G. and Marsh, L.M., 2018. Fra2 Overexpression in Mice Leads to Non-allergic Asthma Development in an IL-13 Dependent Manner. *Frontiers in Immunology*, 9, p.2018. Available from: <https://doi.org/10.3389/fimmu.2018.02018> [Accessed 27 August 2019].
- Gupta, S., Kumar, P., Kaur, H., Sharma, N., Saluja, D., Bharti, A.C. and Das, B.C., 2015. Selective participation of c-Jun with Fra-2/c-Fos promotes aggressive tumor phenotypes and poor prognosis in tongue cancer. *Scientific Reports*, 5, p.16811. Available from: <https://doi.org/10.1038/srep16811> [Accessed 26 September 2016].
- Hagemann, U.B., Mason, J.M., Müller, K.M. and Arndt, K.M., 2008. Selectional and Mutational Scope of Peptides Sequestering the Jun-Fos Coiled-Coil Domain. *Journal of Molecular Biology*, 381(1), pp.73–88. Available from: <https://doi.org/10.1016/j.jmb.2008.04.030> [Accessed 19 September 2015].
- Harbury, P.B., Zhang, T., Kim, P.S. and Alber, T., 1993. A Switch Between Two-, Three-, and Four-Stranded Coiled Coils in GCN4 Leucine Zipper Mutants. *Science*, 262(November), pp.1401–1407. Available from: <https://doi.org/10.1126/science.8248779> [Accessed 2 May 2019].
- Hardie, W.D., Glasser, S.W. and Hagood, J.S., 2009. Emerging concepts in the pathogenesis of lung fibrosis. *American Journal of Pathology*, 175(1), pp.3–16. Available from: <https://doi.org/10.2353/ajpath.2009.081170>.
- Hartmann, M.D., Ridderbusch, O., Zeth, K., Albrecht, R., Testa, O., Woolfson, D.N., Sauer, G., Dunin-Horkawicz, S., Lupas, A.N. and Alvarez, B.H., 2009. A coiled-coil motif that sequesters ions to the hydrophobic core. *Proceedings of the National Academy of Sciences*, 106(40), pp.16950–16955. Available from: <https://doi.org/10.1073/pnas.0907256106> [Accessed 23 September 2016].
- Havranek, J.J. and Harbury, P.B., 2003. Automated design of specificity in molecular recognition. *Nature Structural Biology*, 10(1), pp.45–52. Available from: <https://doi.org/10.1038/nsb877>.
- Hennigan, R.F., Hawker, K.L. and Ozanne, B.W., 1994. Fos-transformation activates genes associated with invasion. *Oncogene*, 9(12), pp.3591–600. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/7970719> [Accessed 28 August 2019].
- Hesketh, J., Campbell, G., Piechaczyk, M. and Blanchard, J.M., 1994. Targeting of c- myc and β -globin coding sequences to cytoskeletal-bound polysomes by c- myc 3' untranslated region. *Biochemical Journal*, 298(1), pp.143–148. Available from: <https://doi.org/10.1042/bj2980143>.

- Hess, J., 2004. AP-1 subunits: quarrel and harmony among siblings. *Journal of Cell Science*, 117(25), pp.5965–5973. Available from: <https://doi.org/10.1242/jcs.01589> [Accessed 24 May 2016].
- Hilfiker-Kleiner, D., Hilfiker, A., Kaminski, K., Schaefer, A., Park, J.K., Michel, K., Quint, A., Yaniv, M., Weitzman, J.B. and Drexler, H., 2005. Lack of JunD promotes pressure overload-induced apoptosis, hypertrophic growth, and angiogenesis in the heart. *Circulation*, 112(10), pp.1470–1477. Available from: <https://doi.org/10.1161/CIRCULATIONAHA.104.518472>.
- Hill, T.A., Shepherd, N.E., Diness, F. and Fairlie, D.P., 2014. Constraining cyclic peptides to mimic protein structure motifs. *Angewandte Chemie - International Edition*, 53(48), pp.13020–13041. Available from: <https://doi.org/10.1002/anie.201401058> [Accessed 26 September 2015].
- Hoang, H.N., Driver, R.W., Beyer, R.L., Hill, T.A., D. de Araujo, A., Plisson, F., Harrison, R.S., Goedecke, L., Shepherd, N.E. and Fairlie, D.P., 2016. Helix Nucleation by the Smallest Known α -Helix in Water. *Angewandte Chemie - International Edition*, 55(29), pp.8275–8279. Available from: <https://doi.org/10.1002/anie.201602079> [Accessed 24 April 2017].
- Hoang, H.N., Song, K., Hill, T.A., Derksen, D.R., Edmonds, D.J., Kok, W.M., Limberakis, C., Liras, S., Loria, P.M., Mascitti, V., Mathiowetz, A.M., Mitchell, J.M., Piotrowski, D.W., Price, D.A., Stanton, R. V., Suen, J.Y., Withka, J.M., Griffith, D.A. and Fairlie, D.P., 2015. Short hydrophobic peptides with cyclic constraints are potent glucagon-like peptide-1 receptor (GLP-1R) agonists. *Journal of Medicinal Chemistry*, 58(9), pp.4080–4085. Available from: <https://doi.org/10.1021/acs.jmedchem.5b00166> [Accessed 24 April 2017].
- Hodges, R.S., Zhou, N.E., Kay, C.M. and Semchuk, P.D., 1990. Synthetic model proteins: contribution of hydrophobic residues and disulfide bonds to protein stability. *Peptide research*, 3(3), pp.123–137.
- Hu, J., Wang, J., Lin, J., Liu, T., Zhong, Y., Liu, J., Zheng, Y., Gao, Y., He, J. and Shang, X., 2019. MD-SVM: a novel SVM-based algorithm for the motif discovery of transcription factor binding sites. *BMC Bioinformatics*, 20(S7), p.200. Available from: <https://doi.org/10.1186/s12859-019-2735-3> [Accessed 21 August 2019].
- Hu, J.C., O'Shea, E.K., Kim, P.S. and Sauer, R.T., 1990. Sequence requirements for coiled-coils: Analysis with λ repressor-GCN4 leucine zipper fusions. *Science*, 250(4986), pp.1400–1403. Available from: <https://doi.org/10.1126/science.2147779>.

- Hu, J.C. and Sauer, R.T., 1992. The Basic-Region Leucine-Zipper Family of DNA Binding Proteins. Springer, Berlin, Heidelberg, pp.82–101. Available from: https://doi.org/10.1007/978-3-642-77356-3_5 [Accessed 28 August 2019].
- Ibrahim, S.A.E.-F., Abudu, A., Jonhson, E., Aftab, N., Conrad, S., Fluck, M., Abd, S., Ibrahim, E.-F., Abudu, A., Jonhson, E., Aftab, N., Conrad, S. and Fluck, M., 2018. The role of AP-1 in self-sufficient proliferation and migration of cancer cells and its potential impact on an autocrine/paracrine loop. *Oncotarget*, 9(76), p.34259. Available from: <https://doi.org/10.18632/ONCOTARGET.26047> [Accessed 24 April 2019].
- Van IJzendoorn, D.G.P., Forghany, Z., Liebelt, F., Vertegaal, A.C., Jochemsen, A.G., Bovée, J.V.M.G., Szuhai, K. and Baker, D.A., 2017. Functional analyses of a human vascular tumor FOS variant identify a novel degradation mechanism and a link to tumorigenesis. *Journal of Biological Chemistry*, 292(52), pp.21282–21290. Available from: <https://doi.org/10.1074/jbc.C117.815845>.
- Ivanov, V.N., Bhoumik, A., Krasilnikov, M., Raz, R., Owen-Schaub, L.B., Levy, D., Horvath, C.M. and Ronai, Z., 2001. Cooperation between STAT3 and c-Jun suppresses Fas transcription. *Molecular Cell*, 7(3), pp.517–528. Available from: [https://doi.org/10.1016/S1097-2765\(01\)00199-X](https://doi.org/10.1016/S1097-2765(01)00199-X).
- Jindrich, K. and Degnan, B.M., 2016. The diversification of the basic leucine zipper family in eukaryotes correlates with the evolution of multicellularity Genome evolution and evolutionary systems biology. *BMC Evolutionary Biology*, 16(1). Available from: <https://doi.org/10.1186/s12862-016-0598-z> [Accessed 27 June 2016].
- Jochum, W., Passequé, E. and Wagner, E.F., 2001. AP-1 in mouse development and tumorigenesis. *Oncogene*, 20(19), pp.2401–2412. Available from: <https://doi.org/10.1038/sj.onc.1204389> [Accessed 25 April 2019].
- Jooss, K.U. and Müller, R., 1995. Deregulation of genes encoding microfilament-associated proteins during Fos-induced morphological transformation. *Oncogene*, 10(3), pp.603–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/7845686> [Accessed 28 August 2019].
- Joung, J.K., Ramm, E.I. and Pabo, C.O., 2000. A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 97(13), pp.7382–7. Available from: <https://doi.org/10.1073/pnas.110149297> [Accessed 19 August 2019].
- Kani, K., Garri, C., Tiemann, K., Malihi, P.D., Punj, V., Nguyen, A.L., Lee, J., Hughes, L.D., Alvarez, R.M., Wood, D.M., Joo, A.Y., Katz, J.E., Agus, D.B. and Mallick, P., 2017.

- JUN-Mediated Downregulation of EGFR Signaling Is Associated with Resistance to Gefitinib in EGFR-mutant NSCLC Cell Lines. *Molecular cancer therapeutics*, 16(8), pp.1645–1657. Available from: <https://doi.org/10.1158/1535-7163.MCT-16-0564> [Accessed 25 April 2019].
- Kaplan, J.B., Reinke, A.W. and Keating, A.E., 2014. Increasing the affinity of selective bZIPbinding peptides through surface residue redesign. *Protein Science*, 23(7), pp.940–953. Available from: <https://doi.org/10.1002/pro.2477> [Accessed 27 June 2016].
- Kappelmann, M., Kuphal, S., Bosserhoff, A., Kuphal, S., Bosserhoff, A., Kuphal, S., Bosserhoff, A. and Kuphal, S., 2014. AP-1/c-Jun transcription factors: Regulation and function in malignant melanoma. *European Journal of Cell Biology*, 93(1–2), pp.76–81. Available from: <https://doi.org/10.1016/j.ejcb.2013.10.003> [Accessed 9 September 2016].
- Kasibhatla, S., Brunner, T., Genestier, L., Echeverri, F., Mahboubi, A. and Green, D.R., 1998. DNA damaging agents induce expression of Fas ligand and subsequent apoptosis in T lymphocytes via the activation of NF- κ B and AP-1. *Molecular Cell*, 1(4), pp.543–551. Available from: [https://doi.org/10.1016/S1097-2765\(00\)80054-4](https://doi.org/10.1016/S1097-2765(00)80054-4).
- Krylov, D., Barchi, J. and Vinson, C., 1998a. Inter-helical interactions in the leucine zipper coiled coil dimer: pH and salt dependence of coupling energy between charged amino acids. *Journal of Molecular Biology*, 279(4), pp.959–972. Available from: <https://doi.org/10.1006/jmbi.1998.1762> [Accessed 23 September 2016].
- Krylov, D., Barchi, J. and Vinson, C., 1998b. Inter-helical interactions in the leucine zipper coiled coil dimer: pH and salt dependence of coupling energy between charged amino acids1. *Journal of Molecular Biology*, 279(4), pp.959–972. Available from: <https://doi.org/10.1006/jmbi.1998.1762> [Accessed 23 September 2016].
- Kuroda, K.O., Meaney, M.J., Uetani, N. and Kato, T., 2008. Neurobehavioral basis of the impaired nurturing in mice lacking the immediate early gene FosB. *Brain Research*, 1211, pp.57–71. Available from: <https://doi.org/10.1016/j.brainres.2008.02.100> [Accessed 12 August 2019].
- Kyriakis, J.M., Banerjee, P., Nikolakaki, E., Dai, T., Rubie, E.A., Ahmad, M.F., Avruch, J. and Woodgett, J.R., 1994. The stress-activated protein kinase subfamily of c-Jun kinases. *Nature*, 369(6476), pp.156–160. Available from: <https://doi.org/10.1038/369156a0> [Accessed 15 May 2019].
- Lamb, P. and McKnight, S.L., 1991. Diversity and specificity in transcriptional regulation:

the benefits of heterotypic dimerization. *Trends in Biochemical Sciences*, 16(November), pp.417–422. Available from: [https://doi.org/10.1016/0968-0004\(91\)90167-T](https://doi.org/10.1016/0968-0004(91)90167-T).

Lamb, R.F., Hennigan, R.F., Turnbull, K., Katsanakis, K.D., MacKenzie, E.D., Birnie, G.D. and Ozanne, B.W., 1997. AP-1-mediated invasion requires increased expression of the hyaluronan receptor CD44. *Molecular and cellular biology*, 17(2), pp.963–76. Available from: <https://doi.org/10.1128/mcb.17.2.963> [Accessed 28 August 2019].

Lappas, M., Riley, C., Lim, R., Barker, G., Rice, G.E., Menon, R. and Permezel, M., 2011. MAPK and AP-1 proteins are increased in term pre-labour fetal membranes overlying the cervix: Regulation of enzymes involved in the degradation of fetal membranes. *Placenta*, 32(12), pp.1016–1025. Available from: <https://doi.org/10.1016/J.PLACENTA.2011.09.011> [Accessed 12 August 2019].

Lathbridge, A. and Mason, J.M., 2018. Computational Competitive and Negative Design to Derive a Specific cJun Antagonist. *Biochemistry*, 57(42), pp.6108–6118. Available from: <https://doi.org/10.1021/acs.biochem.8b00782> [Accessed 7 January 2019].

Lathbridge, A. and Mason, J.M., 2019. Combining Constrained Heptapeptide Cassettes with Computational Design To Create Coiled-Coil Targeting Helical Peptides. *ACS Chemical Biology*, p.acschembio.9b00265. Available from: <https://doi.org/10.1021/acschembio.9b00265> [Accessed 13 June 2019].

Latifa Bakiri, Dominique Lallemand, E.B.-W. and M.Y., Bakiri, L., Lallemand, D., Bossy-Wetzel, E. and Yaniv, M., 2000. Cell cycle-dependent variations in c-Jun and JunB phosphorylation: a role in the control of cyclin D1 expression. *The EMBO Journal*, 19(9), pp.2056–2068. Available from: <https://doi.org/10.1093/emboj/19.9.2056>.

Leach, D.A., Panagopoulos, V., Nash, C., Bevan, C., Thomson, A.A., Selth, L.A. and Buchanan, G., 2017. Cell-lineage specificity and role of AP-1 in the prostate fibroblast androgen receptor cistrome. *Molecular and Cellular Endocrinology*, 439, pp.261–272. Available from: <https://doi.org/10.1016/J.MCE.2016.09.010> [Accessed 25 April 2019].

MacLaren, A., Black, E.J., Clark, W. and Gillespie, D.A.F., 2004. c-Jun-Deficient Cells Undergo Premature Senescence as a Result of Spontaneous DNA Damage Accumulation. *Molecular and Cellular Biology*, 24(20), pp.9006–9018. Available from: <https://doi.org/10.1128/mcb.24.20.9006-9018.2004>.

Majerle, A., Schmieden, D.T., Jerala, R. and Meyer, A.S., 2019. Synthetic Biology for Multiscale Designed Biomimetic Assemblies: From Designed Self-Assembling Biopolymers to Bacterial Bioprinting. *Biochemistry*. Available from:

- Maki, Y., Bos, T.J., Davis, C., Starbuck, M. and Vogt, P.K., 1987. Avian sarcoma virus 17 carries the jun oncogene. *Proceedings of the National Academy of Sciences of the United States of America*, 84(9), pp.2848–52. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/3033666><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC304757>.
- Malorni, L., Giuliano, M., Migliaccio, I., Wang, T., Creighton, C.J., Lupien, M., Fu, X., Hilsenbeck, S.G., Healy, N., De Angelis, C., Mazumdar, A., Trivedi, M. V., Massarweh, S., Gutierrez, C., De Placido, S., Jeselsohn, R., Brown, M., Brown, P.H., Osborne, C.K. and Schiff, R., 2016. Blockade of AP-1 Potentiates Endocrine Therapy and Overcomes Resistance. *Molecular Cancer Research*, 14(5), pp.470–481. Available from: <https://doi.org/10.1158/1541-7786.MCR-15-0423> [Accessed 28 August 2019].
- Marconcini, L., Marchio, S., Morbidelli, L., Cartocci, E., Albini, A., Ziche, M., Bussolino, F. and Oliviero, S., 1999. c-fos-induced growth factor/vascular endothelial growth factor D induces angiogenesis in vivo and in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, 96(17), pp.9671–6. Available from: <https://doi.org/10.1073/pnas.96.17.9671> [Accessed 28 August 2019].
- Mason, J.M., 2010. Design and development of peptides and peptide mimetics as antagonists for therapeutic intervention. *Future Medicinal Chemistry*, 2(12), pp.1813–1822. Available from: <https://doi.org/10.4155/fmc.10.259> [Accessed 19 September 2015].
- Mason, J.M. and Arndt, K.M., 2004. Coiled coil domains: Stability, specificity, and biological implications. *ChemBioChem*, 5(2), pp.170–176. Available from: <https://doi.org/10.1002/cbic.200300781> [Accessed 19 September 2015].
- Mason, J.M., Hagemann, U.B. and Arndt, K.M., 2007. Improved stability of the Jun-Fos activator protein-1 coiled coil motif: A stopped-flow circular dichroism kinetic analysis. *Journal of Biological Chemistry*, 282(32), pp.23015–23024. Available from: <https://doi.org/10.1074/jbc.M701828200> [Accessed 19 September 2015].
- Mason, J.M., Müller, K.M. and Arndt, K.M., 2007. Positive aspects of negative design: Simultaneous selection of specificity and interaction stability. *Biochemistry*, 46(16), pp.4804–4814. Available from: <https://doi.org/10.1021/bi602506p> [Accessed 19 September 2015].
- Mason, J.M., Schmitz, M.A., Muller, K.M. and Arndt, K.M., 2006. Semirational design of Jun-Fos coiled coils with increased affinity: Universal implications for leucine zipper

- prediction and design. *Proceedings of the National Academy of Sciences*, 103(24), pp.8989–8994. Available from: <https://doi.org/10.1073/pnas.0509880103> [Accessed 19 September 2015].
- Mathonet, P., Ioannou, A., Betley, J. and Ullman, C., 2011. CIS display, a DNA-based in vitro selection technology for therapeutic peptides. *Chimica Oggi-Chemistry Today*, 29(2), pp.10–12.
- Miao, F., Zhang, M., Zhao, Y., Li, X., Yao, R., Wu, F., Huang, R., Li, K., Miao, S., Ma, C., Ju, H., Song, W. and Wang, L., 2017. RHBDD1 upregulates EGFR via the AP-1 pathway in colorectal cancer. *Oncotarget*, 8(15), pp.25251–25260. Available from: <https://doi.org/10.18632/oncotarget.15694> [Accessed 8 January 2018].
- Michnick, S.W., Remy, I., Campbell-Valois, F.X., Vallee-Belisle, A. and Pelletier, J.N., 2000. Detection of protein-protein interactions by protein fragment complementation strategies. *Methods in Enzymology*, 328, pp.208–230.
- Milde-Langosch, K., 2005. The Fos family of transcription factors and their role in tumourigenesis. *European Journal of Cancer*, 41(16), pp.2449–2461. Available from: <https://doi.org/10.1016/j.ejca.2005.08.008> [Accessed 26 September 2016].
- Monera, O.D., Zhou, N.E., Kay, C.M. and Hodges, R.S., 1993. Comparison of antiparallel and parallel two-stranded alpha-helical coiled-coils. Design, synthesis, and characterization. *The Journal of biological chemistry*, 268(26), pp.19218–27. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/8366074>.
- Mukherjee, A., Hollern, D.P., Williams, O.G., Rayburn, T.S., Byrd, W.A., Yates, C. and Jones, J.D., 2018. A Review of FOXI3 Regulation of Development and Possible Roles in Cancer Progression and Metastasis. *Frontiers in Cell and Developmental Biology*, 6(July), pp.1–9. Available from: <https://doi.org/10.3389/fcell.2018.00069>.
- Nautiyal, S., Woolfson, D.N., King, D.S. and Alber, T., 1995. A Designed Heterotrimeric Coiled Coil. *Biochemistry*, 34(37), pp.11645–11651. Available from: <https://doi.org/10.1021/bi00037a001> [Accessed 7 May 2019].
- O’Neil, K.T. and DeGrado, W.F., 1990. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science*, 250(4981), pp.646–651. Available from: <https://doi.org/10.1126/science.2237415>.
- O’Shea, E.K., Klemm, J.D., Kim, P.S. and Alber, T., 1991. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science*, 254(5031), pp.539–544. Available from: <https://doi.org/10.1126/science.1948029>.

- O'Shea, E.K., Lumb, K.J. and Kim, P.S., 1993. Peptide 'Velcro': Design of a heterodimeric coiled coil. *Current Biology*, 3(10), pp.658–667. Available from: [https://doi.org/10.1016/0960-9822\(93\)90063-T](https://doi.org/10.1016/0960-9822(93)90063-T) [Accessed 26 May 2016].
- Odegrip, R., Coomber, D., Eldridge, B., Hederer, R., Kuhlman, P.A., Ullman, C., FitzGerald, K. and McGregor, D., 2004. CIS display: In vitro selection of peptides from libraries of protein-DNA complexes. *Proceedings of the National Academy of Sciences*, 101(9), pp.2806–2810. Available from: <https://doi.org/10.1073/pnas.0400219101>.
- Oien, K.A., Vass, J Keith, Downie, I., Fullarton, G., Keith, W.N., Dennis, J.L., Wit, E.C., Gaffney, D., Pullinger, C.R., O'Reilly, D.S.J., Hoffs, M.S., Cameron, I., Kulkarni, M. V, Kane, J.P., Schumaker, V.N., Watts, G.F., Packard, C.J., Johnston, I.M., Spence, H.J., Winnie, J.N., McGarry, L., Vass, J K, Meagher, L., Stapleton, G., Ozanne, B.W., Hawker, K.L., Iturralde, M., Oria, R., Brock, J.H., McGregor, S.J., Naves, M.L., Neill, R., Jamieson, T. and Birnie, G.D., 2000. Regulation of a multigenic invasion programme by the transcription factor, AP-1: re-expression of a down-regulated gene, TSC-36, inhibits invasion. *Oncogene*, 19(47), pp.5348–58. Available from: <https://doi.org/10.1038/sj.onc.1203927>.
- Papavassiliou, K.A. and Papavassiliou, A.G., 2016. Transcription Factor Drug Targets. *Journal of Cellular Biochemistry*, 117(12), pp.2693–2696. Available from: <https://doi.org/10.1002/jcb.25605> [Accessed 23 June 2017].
- Park, J.M., Adam, R.M., Peters, C.A., Guthrie, P.D., Sun, Z., Klagsbrun, M. and Freeman, M.R., 1999. AP-1 mediates stretch-induced expression of HB-EGF in bladder smooth muscle cells. *American Journal of Physiology - Cell Physiology*, 277(2 46-2).
- Parry, D.A., 1982. Coiled-coils in alpha-helix-containing proteins: analysis of the residue types within the heptad repeat and the use of these data in the prediction of coiled-coils in other proteins. *Bioscience reports*, 2(12), pp.1017–24. Available from: <https://doi.org/10.1007/bf01122170> [Accessed 21 August 2019].
- Parry, D.A.D., Crewther, W.G., Fraser, R.D.B. and MacRae, T.P., 1977. Structure of α -keratin: Structural implication of the amino acid sequences of the type I and type II chain segments. *Journal of Molecular Biology*, 113(2), pp.449–454. Available from: [https://doi.org/10.1016/0022-2836\(77\)90153-X](https://doi.org/10.1016/0022-2836(77)90153-X) [Accessed 21 August 2019].
- Passegué, E. and Wagner, E.F., 2000. JunB suppresses cell proliferation by transcriptional activation of p16(INK4a) expression. *The EMBO journal*, 19(12), pp.2969–79. Available from: <https://doi.org/10.1093/emboj/19.12.2969> [Accessed 15 May 2019].
- Patgiri, A., Jochim, A.L. and Arora, P.S., 2008. A Hydrogen Bond Surrogate Approach for

- Stabilization of Short Peptide Sequences in α -Helical Conformation. *Accounts of Chemical Research*, 41(10), pp.1289–1300. Available from: <https://doi.org/10.1021/ar700264k> [Accessed 4 April 2019].
- Pelletier, J.N., Campbell-Valois, F.-X. and Michnick, S.W., 1998. Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. *Proceedings of the National Academy of Sciences*, 95(21), pp.12141–12146. Available from: <https://doi.org/10.1073/pnas.95.21.12141>.
- Pollok, B.A. and Heim, R., 1999. Using GFP in FRET-based applications. *Trends in cell biology*, 9(2), pp.57–60. Available from: [https://doi.org/10.1016/S0962-8924\(98\)01434-2](https://doi.org/10.1016/S0962-8924(98)01434-2) [Accessed 19 August 2019].
- Potapov, V., Kaplan, J.B. and Keating, A.E., 2015. Data-Driven Prediction and Design of bZIP Coiled-Coil Interactions. *PLoS Computational Biology*, 11(2), p.e1004046. Available from: <https://doi.org/10.1371/journal.pcbi.1004046> [Accessed 27 June 2016].
- Rajasekaran, S., Vaz, M. and Reddy, S.P., 2012. Fra-1/AP-1 transcription factor negatively regulates pulmonary fibrosis in vivo. *PLoS ONE*, 7(7). Available from: <https://doi.org/10.1371/journal.pone.0041611>.
- Ransone, L.J., Kerr, L.D., Schmitt, M.J., Wamsley, P. and Verma, I.M., 1993. The bZIP domains of Fos and Jun mediate a physical association with the TATA box-binding protein. *Gene expression*, 3(1), pp.37–48. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/7685215> [Accessed 25 April 2019].
- Rao, T., Ruiz-Gómez, G., Hill, T.A., Hoang, H.N., Fairlie, D.P. and Mason, J.M., 2013. Truncated and Helix-Constrained Peptides with High Affinity and Specificity for the cFos Coiled-Coil of AP-1. *PLoS ONE*, 8(3), p.e59415. Available from: <https://doi.org/10.1371/journal.pone.0059415> [Accessed 26 September 2016].
- Rastogi, S., Shukla, S., Kalaivani, M. and Singh, G.N., 2019. Peptide-based therapeutics: quality specifications, regulatory considerations, and prospects. *Drug Discovery Today*, 24(1), pp.148–162. Available from: <https://doi.org/10.1016/j.drudis.2018.10.002>.
- Rebollo, A., Dumoutier, L., Renauld, J.-C., Zaballo, A., Ayllon, V. and Martinez-A., C., 2000. Bcl-3 Expression Promotes Cell Survival following Interleukin-4 Deprivation and Is Controlled by AP1 and AP1-Like Transcription Factors. *Molecular and Cellular Biology*, 20(10), pp.3407–3416. Available from: <https://doi.org/10.1128/mcb.20.10.3407-3416.2000>.

- Reinke, A.W., Baek, J., Ashenberg, O. and Keating, A.E., 2013. Networks of bZIP protein-protein interactions diversified over a billion years of evolution. *Science*, 340(6133), pp.730–734. Available from: <https://doi.org/10.1126/science.1233465> [Accessed 27 June 2016].
- Reinke, A.W., Grigoryan, G. and Keating, A.E., 2010. Identification of bZIP interaction partners of viral proteins HBZ, MEQ, BZLF1, and K-bZIP using coiled-coil arrays. *Biochemistry*, 49(9), pp.1985–1997. Available from: <https://doi.org/10.1021/bi902065k> [Accessed 27 June 2016].
- Remy, I., Galarneau, A. and Michnick, S.W., 2002. Detection and visualization of protein interactions with protein fragment complementation assays. *Methods in molecular biology (Clifton, N.J.)*, 185(5), pp.447–459.
- Remy, I. and Michnick, S.W., 2015. Mapping biochemical networks with protein fragment complementation assays. *Protein-Protein Interactions: Methods and Applications: Second Edition*, 261, pp.467–481. Available from: https://doi.org/10.1007/978-1-4939-2425-7_31.
- Rezaei Araghi, R., Bird, G.H., Ryan, J.A., Jenson, J.M., Godes, M., Pritz, J.R., Grant, R.A., Letai, A., Walensky, L.D. and Keating, A.E., 2018. Iterative optimization yields Mcl-1-targeting stapled peptides with selective cytotoxicity to Mcl-1-dependent cancer cells. *Proceedings of the National Academy of Sciences of the United States of America*, 115(5), pp.E886–E895. Available from: <https://doi.org/10.1073/pnas.1712952115>.
- Ricci, R., Eriksson, U., Oudit, G.Y., Eferl, R., Akhmedov, A., Sumara, I., Sumara, G., Kassiri, Z., David, J.P., Bakiri, L., Sasse, B., Idarraga, M.H., Rath, M., Kurz, D., Theussl, H.C., Perriard, J.C., Backx, P., Penninger, J.M. and Wagner, E.F., 2005. Distinct functions of junD in cardiac hypertrophy and heart failure. *Genes and Development*, 19(2), pp.208–213. Available from: <https://doi.org/10.1101/gad.327005>.
- Rincon, M. and Flavell, R.A., 1994. *AP-1 transcriptional activity requires both T-cell receptor-mediated and co-stimulatory signals in primary T lymphocytes*. European Molecular Biology Organization. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC395364/pdf/emboj00066-0188.pdf> [Accessed 13 May 2019].
- Rodríguez-Martínez, J.A., Reinke, A.W., Bhimsaria, D., Keating, A.E. and Ansari, A.Z., 2017. Combinatorial bZIP dimers display complex DNA-binding specificity landscapes. *eLife*, 6, p.e19272. Available from: <https://doi.org/10.7554/eLife.19272> [Accessed 13 February 2017].

- Sato, H. and Seiki, M., 1993. Regulatory mechanism of 92 kDa type IV collagenase gene expression which is associated with invasiveness of tumor cells. *Oncogene*, 8(2), pp.395–405. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/8426746> [Accessed 12 August 2019].
- Schreiber, M., Kolbus, A., Piu, F., Szabowski, A., Möhle-Steinlein, U., Tian, J., Karin, M., Angel, P. and Wagner, E.F., 1999. Control of cell cycle progression by c-Jun is p53 dependent. *Genes and Development*, 13(5), pp.607–619. Available from: <https://doi.org/10.1101/gad.13.5.607>.
- Schreiber, M., Wang, Z.Q., Jochum, W., Fetka, I., Elliott, C. and Wagner, E.F., 2000. Placental vascularisation requires the AP-1 component Fra1. *Development*, 127(22), pp.4937–4948. Available from: <http://dev.biologists.org/content/127/22/4937.abstract>.
- Seldeen, K.L., McDonald, C.B., Deegan, B.J., Bhat, V. and Farooq, A., 2010. Dissecting the role of leucine zippers in the binding of bZIP domains of Jun transcription factor to DNA. *Biochemical and Biophysical Research Communications*, 394(4), pp.1030–1035. Available from: <https://doi.org/10.1016/j.bbrc.2010.03.116> [Accessed 27 June 2016].
- Shameer, K., Pugalenth, G., Kandaswamy, K.K., Suganthan, P.N., Archunan, G. and Sowdhamini, R., 2010. Bioinformatics and Biology Insights Insights into protein sequence and structure-Derived Features Mediating 3D Domain swapping Mechanism using support Vector Machine Based Approach. *Bioinformatics and Biology Insights*, 4, pp.33–42. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2901629/> [Accessed 18 January 2017].
- Shaulian, E., 2010. AP-1 — The Jun proteins: Oncogenes or tumor suppressors in disguise? *Cellular Signalling*, 22(6), pp.894–899. Available from: <https://doi.org/10.1016/J.CELLSIG.2009.12.008> [Accessed 24 April 2019].
- Shlizerman, C., Atanassov, A., Berkovich, I., Ashkenasy, G. and Ashkenasy, N., 2010. De novo designed coiled-coil proteins with variable conformations as components of molecular electronic devices. *Journal of the American Chemical Society*, 132(14), pp.5070–5076. Available from: <https://doi.org/10.1021/ja907902h>.
- Singh, M. and Kim, P.S., 2001. Towards predicting coiled-coil protein interactions. *Proceedings of the Annual International Conference on Computational Molecular Biology, RECOMB*, pp.279–286.
- Sitras, V., Paulssen, R.H., Grønnaas, H., Vårtun, Å. and Acharya, G., 2008. Gene expression profile in labouring and non-labouring human placenta near term. *Molecular Human*

- Staber, P.B., Vesely, P., Haq, N., Ott, R.G., Funato, K., Bambach, I., Fuchs, C., Schauer, S., Linkesch, W., Hrzenjak, A., Dirks, W.G., Sexl, V., Bergler, H., Kadin, M.E., Sternberg, D.W., Kenner, L. and Hoefler, G., 2007. The oncoprotein NPM-ALK of anaplastic large-cell lymphoma induces JUNB transcription via ERK1/2 and JunB translation via mTOR signaling. *Blood*, 110(9), pp.3374–3383. Available from: <https://doi.org/10.1182/blood-2007-02-071258>.
- Sun, D., Zhang, M., Li, Y., Mei, S., Qin, J. and Yan, J., 2019. c-Jun/AP-1 is upregulated in an Ang II-induced abdominal aortic aneurysm formation model and mediates Chop expression in mouse aortic smooth muscle cells. *Molecular Medicine Reports*, 49(5), pp.3459–3468. Available from: <https://doi.org/10.3892/mmr.2019.10017>.
- Szalóki, N., Krieger, J.W., Komáromi, I., Tóth, K. and Vámosi, G., 2015. Evidence for Homodimerization of the c-Fos Transcription Factor in Live Cells Revealed by Fluorescence Microscopy and Computer Modeling. *Molecular and cellular biology*, 35(21), pp.3785–98. Available from: <https://doi.org/10.1128/MCB.00346-15> [Accessed 19 August 2019].
- Tapeinou, A., Matsoukas, M.T., Simal, C. and Tselios, T., 2015. Review cyclic peptides on a merry-go-round; towards drug design. *Biopolymers*, 104(5), pp.453–461. Available from: <https://doi.org/10.1002/bip.22669>.
- Taylor, J.W., 2002. The synthesis and study of side-chain lactam-bridged peptides. *Biopolymers - Peptide Science Section*, 66(1), pp.49–75. Available from: <https://doi.org/10.1002/bip.10203> [Accessed 26 November 2018].
- Tewari, D., Nabavi, S.F., Nabavi, S.M., Sureda, A., Farooqi, A.A., Atanasov, A.G., Vacca, R.A., Sethi, G. and Bishayee, A., 2018. Targeting activator protein 1 signaling pathway by bioactive natural agents: Possible therapeutic strategy for cancer prevention and intervention. *Pharmacological Research*, 128, pp.366–375. Available from: <https://doi.org/10.1016/J.PHRS.2017.09.014> [Accessed 24 April 2019].
- Thansandote, P., Harris, R.M., Dexter, H.L., Simpson, G.L., Pal, S., Upton, R.J. and Valko, K., 2015. Improving the passive permeability of macrocyclic peptides: Balancing permeability with other physicochemical properties. *Bioorganic and Medicinal Chemistry*, 23(2), pp.322–327. Available from: <https://doi.org/10.1016/j.bmc.2014.11.034>.
- Thomas, F., Niitsu, A., Oregioni, A., Bartlett, G.J. and Woolfson, D.N., 2017. Conformational

- Dynamics of Asparagine at Coiled-Coil Interfaces. *Biochemistry*, 56(50), pp.6544–6554. Available from: <https://doi.org/10.1021/acs.biochem.7b00848> [Accessed 10 July 2018].
- Trop-Steinberg, S. and Azar, Y., 2017. AP-1 Expression and its Clinical Relevance in Immune Disorders and Cancer. *The American Journal of the Medical Sciences*, 353(5), pp.474–483. Available from: <https://doi.org/10.1016/J.AMJMS.2017.01.019> [Accessed 25 April 2019].
- Tyagi, A., Vishnoi, K., Kaur, H., Srivastava, Y., Roy, B.G., Das, B.C. and Bharti, A.C., 2017. Cervical cancer stem cells manifest radioresistance: Association with upregulated AP-1 activity. *Scientific Reports*, 7(1), p.4781. Available from: <https://doi.org/10.1038/s41598-017-05162-x> [Accessed 8 January 2018].
- Ucero, A.C., Bakiri, L., Roediger, B., Suzuki, M., Jimenez, M., Mandal, P., Braghetta, P., Bonaldo, P., Paz-Ares, L., Fustero-Torre, C., Ximenez-Embun, P., Hernandez, A.I., Megias, D. and Wagner, E.F., 2019. Fra-2–expressing macrophages promote lung fibrosis. *The Journal of Clinical Investigation*, 129(8), pp.3293–3309. Available from: <https://doi.org/10.1172/JCI125366> [Accessed 28 August 2019].
- Uchihashi, S., Fukumoto, H., Onoda, M., Hayakawa, H., Ikushiro, S. and Sakaki, T., 2011. Metabolism of the c-Fos/activator protein-1 inhibitor T-5224 by multiple human UDP-glucuronosyltransferase isoforms. *Drug Metabolism and Disposition*, 39(5), pp.803–813. Available from: <https://doi.org/10.1124/dmd.110.037952>.
- Udayappan, U.K. and Casey, P.J., 2017. c-Jun Contributes to Transcriptional Control of GNA12 Expression in Prostate Cancer Cells. *Molecules (Basel, Switzerland)*, 22(4). Available from: <https://doi.org/10.3390/molecules22040612> [Accessed 25 April 2019].
- Uluçkan, Ö. and Wagner, E.F., 2016. Role of IL-17A signalling in psoriasis and associated bone loss. *Clinical and Experimental Rheumatology*, 34(4 Suppl 98), pp.S17–S20. Available from: <http://www.clinexprheumatol.org/article.asp?a=10853> [Accessed 26 September 2016].
- Vagner, J., Qu, H. and Hruby, V.J., 2008. Peptidomimetics, a synthetic tool of drug discovery. *Current Opinion in Chemical Biology*, 12(3), pp.292–296. Available from: <https://doi.org/10.1016/j.cbpa.2008.03.009>.
- Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A. and Luscombe, N.M., 2009. A census of human transcription factors: Function, expression and evolution. *Nature Reviews Genetics*, 10(4), pp.252–263. Available from: <https://doi.org/10.1038/nrg2538>.

- Ventura, J.-J., Kennedy, N.J., Lamb, J.A., Flavell, R.A. and Davis, R.J., 2003. c-Jun NH2-Terminal Kinase Is Essential for the Regulation of AP-1 by Tumor Necrosis Factor. *Molecular and Cellular Biology*, 23(8), pp.2871–2882. Available from: <https://doi.org/10.1128/MCB.23.8.2871-2882.2003>.
- Wagner, E.F., 2000. JunB suppresses cell proliferation by transcriptional activation of p16 INK4a expression. *The EMBO Journal*, 19(12), pp.2969–2979.
- Walensky, L.D. and Bird, G.H., 2015. Hydrocarbon-Stapled Peptides: Principles, Practice, and Progress. *Journal of Medicinal Chemistry*, 57(15), pp.6275–6288. Available from: <https://doi.org/10.1021/jm4011675> [Accessed 26 November 2018].
- Walshaw, J. and Woolfson, D.N., 2001. SOCKET: A program for identifying and analysing coiled-coil motifs within protein structures. *Journal of Molecular Biology*, 307(5), pp.1427–1450. Available from: <https://doi.org/10.1006/jmbi.2001.4545>.
- Wang, K., Jin, S., Fan, D., Wang, M., Xing, N. and Niu, Y., 2017. Anti-proliferative activities of finasteride in benign prostate epithelial cells require stromal fibroblasts and c-Jun gene. *PLoS ONE*, 12(2), pp.1–12. Available from: <https://doi.org/10.1371/journal.pone.0172233> [Accessed 25 April 2019].
- Wang, Z., Liang, J., Schellander, K., Wang, Z., Liang, J., Schellander, K., Wagner, E.F. and Gngoriadis, A.E., 1995. c- fos -induced Osteosarcoma Formation in Transgenic Mice : Cooperativity with c- jun and the Role of Endogenous c- fos. *Cancer Res*, 55(24), pp.6244–6251. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/8521421>.
- Watanabe, H., Saitoh, K., Kameda, T., Murakami, M., Niikura, Y., Okazaki, S., Morishita, Y., Mori, S., Yokouchi, Y., Kuroiwa, A. and Iba, H., 2002. Chondrocytes as a specific target of ectopic Fos expression in early development. *Proceedings of the National Academy of Sciences*, 94(8), pp.3994–3999. Available from: <https://doi.org/10.1073/pnas.94.8.3994>.
- Whitfield, J., Neame, S.J., Paquet, L., Bernard, O. and Ham, J., 2001. Dominant-negative c-Jun promotes neuronal survival by reducing BIM expression and inhibiting mitochondrial cytochrome c release. *Neuron*, 29(3), pp.629–643. Available from: [https://doi.org/10.1016/S0896-6273\(01\)00239-2](https://doi.org/10.1016/S0896-6273(01)00239-2).
- Williams, R.W., Chang, A., Juretić, D. and Loughran, S., 1987. Secondary structure predictions and medium range interactions. *Biochimica et Biophysica Acta (BBA)/Protein Structure and Molecular*, 916(2), pp.200–204. Available from: [https://doi.org/10.1016/0167-4838\(87\)90109-9](https://doi.org/10.1016/0167-4838(87)90109-9) [Accessed 23 September 2016].
- Wood, C.W. and Woolfson, D.N., 2018. CCBUILDER 2.0: Powerful and accessible coiled-coil

- modeling. *Protein Science*, 27(1), pp.103–111. Available from: <https://doi.org/10.1002/pro.3279> [Accessed 8 January 2018].
- Woolfson, D.N. and Alber, T., 1995. Predicting oligomerization states of coiled coils. *Protein Science*, 4(8), pp.1596–1607. Available from: <https://doi.org/10.1021/bi00037a001> [Accessed 20 August 2019].
- Xanthoudakis, S., Miao, G.G., Curran, T., Droge, W., Schulze-Osthoff, K., Henderson, W.R. and Kahn, M., 1994. The redox and DNA-repair activities of Ref-1 are encoded by nonoverlapping domains. *Proceedings of the National Academy of Sciences of the United States of America*, 91(1), pp.23–7. Available from: <https://doi.org/10.1073/pnas.91.1.23> [Accessed 27 August 2019].
- Ye, N., Ding, Y., Wild, C., Shen, Q. and Zhou, J., 2014. Small molecule inhibitors targeting activator protein 1 (AP-1). *Journal of Medicinal Chemistry*, 57(16), pp.6930–6948. Available from: <https://doi.org/10.1021/jm5004733> [Accessed 23 June 2017].
- Yeates, T.O., Liu, Y. and Laniado, J., 2016. The design of symmetric protein nanomaterials comes of age in theory and practice. *Current Opinion in Structural Biology*, 39, pp.134–143. Available from: <https://doi.org/10.1016/J.SBI.2016.07.003> [Accessed 28 August 2019].
- Yeh, J.E., Toniolo, P.A. and Frank, D.A., 2013. Targeting transcription factors: Promising new strategies for cancer therapy. *Current Opinion in Oncology*, 25(6), pp.652–658. Available from: <https://doi.org/10.1097/01.cco.0000432528.88101.1a> [Accessed 23 June 2017].
- Yin, Z., Machius, M., Nestler, E.J. and Rudenko, G., 2017. Activator protein-1: Redox switch controlling structure and DNA-binding. *Nucleic Acids Research*, 45(19), pp.11425–11436. Available from: <https://doi.org/10.1093/nar/gkx795> [Accessed 25 April 2019].
- Yonemoto, H., Young, C.B., Ross, J.T., Guilbert, L.L., Fairclough, R.J. and Olson, D.M., 2006. Changes in Matrix Metalloproteinase (MMP)-2 and MMP-9 in the Fetal Amnion and Chorion During Gestation and at Term and Preterm Labor. *Placenta*, 27(6–7), pp.669–677. Available from: <https://doi.org/10.1016/j.placenta.2005.05.014> [Accessed 12 August 2019].
- Young, K.H., 1998. Yeast Two-hybrid: So Many Interactions, (in) So Little Time.... *Biology of Reproduction*, 58(2), pp.302–311. Available from: <https://doi.org/10.1095/biolreprod58.2.302> [Accessed 19 August 2019].
- Zeng, F. and Zimmerman, S.C., 1997. Dendrimers in Supramolecular Chemistry: From Molecular Recognition to Self-Assembly. *Chemical Reviews*, 97(5), pp.1681–1712.

Available from: <https://doi.org/10.1021/cr9603892>.

Zenz, R., Eferl, R., Scheinecker, C., Redlich, K., Smolen, J., Schonthaler, H.B., Kenner, L., Tschachler, E. and Wagner, E.F., 2007. Activator protein 1 (Fos/Jun) functions in inflammatory bone and skin disease. *Arthritis Research & Therapy*, 10(1), p.201. Available from: <https://doi.org/10.1186/ar2338> [Accessed 27 August 2019].

Zhou, F., Grigoryan, G., Lustig, S.R., Keating, A.E., Ceder, G. and Morgan, D., 2005. Coarse-graining protein energetics in sequence variables. *Physical Review Letters*, 95(14), pp.12–15. Available from: <https://doi.org/10.1103/PhysRevLett.95.148103>.

Zuo, Z., Gandhi, N.S., Arndt, K.M. and Mancera, R.L., 2012. Free energy calculations of the interactions of c-Jun-based synthetic peptides with the c-Fos protein. *Biopolymers*, 97(11), pp.899–909. Available from: <https://doi.org/10.1002/bip.22099> [Accessed 24 May 2016].

PAGE INTENTIONALLY LEFT BLANK

APPENDIX CHAPTER FOLLOWS

CHAPTER A1 - COMPUTATIONAL PREDICTION AND DESIGN FOR CREATING ITERATIVELY LARGER HETEROSPECIFIC COILED COIL SETS


Reproduced with permission from:

Crooks, R. O., **Lathbridge, A.**, Panek, A. S., and Mason, J. M. (2017) Computational Prediction and Design for Creating Iteratively Larger Heterospecific Coiled Coil Sets.

Biochemistry 56, 1573–1584

Copyright 2019 American Chemical Society.

Appendix 6B: Statement of Authorship

This declaration concerns the article entitled:			
Computational Prediction and Design for Creating Iteratively Larger Heterospecific Coiled Coil Sets			
Publication status (tick one)			
Draft manuscript <input type="checkbox"/> Submitted <input type="checkbox"/> In review <input type="checkbox"/> Accepted <input type="checkbox"/> Published <input checked="" type="checkbox"/>			
Publication details (reference)	Crooks, R. O., Lathbridge, A., Panek, A. S., and Mason, J. M. (2017) Computational Prediction and Design for Creating Iteratively Larger Heterospecific Coiled Coil Sets. <i>Biochemistry</i> 56, 1573–1584		
Copyright status (tick the appropriate statement)			
I hold the copyright for this material <input type="checkbox"/> Copyright is retained by the publisher, but I have been given permission to replicate the material here <input checked="" type="checkbox"/>			
Candidate's contribution to the paper (provide details, and also indicate as a percentage)	<p>The candidate contributed to / <u>considerably contributed</u> to / predominantly executed the...</p> <p>Formulation of ideas:</p> <p>20% - Contributed <i>in-silico</i> ideas</p> <p>Design of methodology:</p> <p>50% - Developed software</p> <p>Experimental work:</p> <p>50% - Screening and data analysis</p> <p>Presentation of data in journal format:</p> <p>50% - Contributed to manuscript</p>		
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.		
Signed		Date	10/06/19

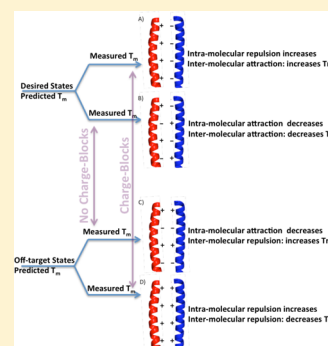
Computational Prediction and Design for Creating Iteratively Larger Heterospecific Coiled Coil Sets

Richard O. Crooks, Alexander Lathbridge, Anna S. Panek, and Jody M. Mason*

Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, U.K.

S Supporting Information

ABSTRACT: A major biochemical goal is the ability to mimic nature in engineering highly specific protein–protein interactions (PPIs). We previously devised a computational interactome screen to identify eight peptides that form four heterospecific dimers despite 32 potential off-targets. To expand the speed and utility of our approach and the PPI toolkit, we have developed new software to derive much larger heterospecific sets (≥ 24 peptides) while directing against antiparallel off-targets. It works by predicting T_m values for every dimer on the basis of core, electrostatic, and helical propensity components. These guide interaction specificity, allowing heterospecific coiled coil (CC) sets to be incrementally assembled. Prediction accuracy is experimentally validated using circular dichroism and size exclusion chromatography. Thermal denaturation data from a 22-CC training set were used to improve software prediction accuracy and verified using a 136-CC test set consisting of eight predicted heterospecific dimers and 128 off-targets. The resulting software, qCIPA, individually now weighs core $a\text{-}a'$ (II/NN/NI) and electrostatic $g\text{-}e^{+1}$ (EE/EK/KK) components. The expanded data set has resulted in emerging sequence context rules for otherwise energetically equivalent CCs; for example, introducing intrahelical electrostatic charge blocks generated increased stability for designed CCs while concomitantly decreasing the stability of off-target CCs. Coupled with increased prediction accuracy and speed, the approach can be applied to a wide range of downstream chemical and synthetic biology applications, in addition more generally to impose specificity on structurally unrelated PPIs.



Protein structures and their interactions form via complex arrangements of cooperative interactions, making *de novo* design of heterospecific protein–protein interactions (PPIs) very difficult. There is a great shortage in the number PPI components that are increasingly needed in biological applications, where the specificity of interaction is important and where large numbers of heterospecific peptide pairs would be beneficial.^{1,2} For example, this unmet need in protein science includes applications in peptide labeling (e.g., monitoring biochemical processes without the need for large tags such as GFP), in delivery of drugs or toxins, in protein purification and labeling applications as high-specificity affinity tags, in creation of large nanostructures such as tetrahedral cages or conductive nanowires, in biomaterials such as reversible hydrogels that assemble or disassemble according to pH or temperature change, and in disease modulation, and many other uses as specific cognate pairs in the synthetic biology toolkit.^{3–5} Specificity of protein–protein interaction and recognition is also essential for normal physiology, with protein interaction network imbalances associated with a wide range of diseases. A major drawback in applying proteins and peptides to such applications is the limited number of exquisitely specific orthogonal PPI-forming peptides that are available. This is because, despite considerable effort, sequence to structure information relating to the protein folding problem is largely unavailable.⁶ However, this is becoming possible for systems such as coiled coils (CCs), where the rules translating how primary sequence dictates quaternary structure are

becoming increasingly understood.^{7,8} The CC motif is an interesting PPI model as it is a simplistic example of quaternary structure and commonly found in a wide range of therapeutically relevant proteins. Utilizing CCs as a model to predict the stability and specificity of protein dimerization directly from the primary sequence is therefore an important and tractable goal. This is because despite the apparent simplicity, CCs are highly specific in the interactions that they drive. Using knowledge of this type of protein fold, we have used *de novo* design to generate the formation of specific CCs that can be applied in a wide range of applications. Here we utilize newly created software to allow a great expansion of the number of specific CC-forming peptides and produce large customized sets of peptides that vary according to the needs of the user. To meet these goals, we have built and tested freely available computational tools (see the [Supporting Information](#)) that allow the user to derive large numbers of structurally similar orthogonal pairs with the potential to create excellent candidates for scaffold parts.

For the design of coiled coil pairs, although a good qualitative understanding of sequences that form a parallel dimeric CC exists, a quantitative understanding of how precise residue placements within dictate both stability and specificity is still

Received: January 19, 2017

Revised: March 3, 2017

Published: March 7, 2017

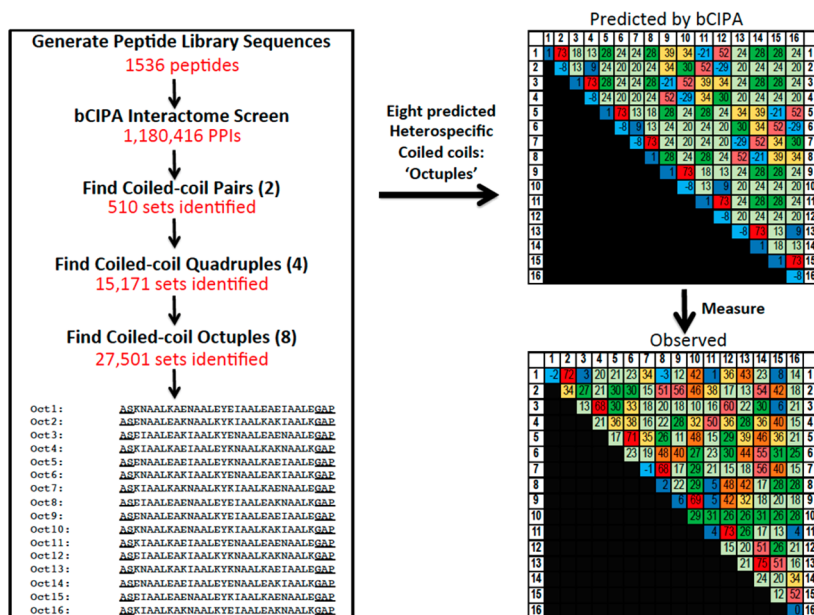


Figure 1. Software for computationally guiding the derivation of specific coiled coils. First, “Generate Library” was used to create a complete list of the peptide library (in this case, a 4096-member peptide library was reduced to 1536 members). Next, the “bCIPA Interactome Screen” was used to predict the T_m of each potential CC within the 1180416-pairwise CC interactome. Next “Find Pairs” was used to identify sets of peptides that, according to the criteria input by the user, are predicted to be heterospecific when combined. In this case, 510 such sets were identified. Next “Find Quadruples” was used to combine pairs of CCs to identify groups of four CCs that are predicted to be heterospecific when combined. Here, 15171 sets were identified. Finally, “Find Octuples” combined sets of quadruples to identify 27501 sets of 16 peptides that are predicted to be heterospecific when mixed. The 16-peptide set used in this study is shown, with additional capping sequences underlined. Shown on the right are the bCIPA-predicted and measured thermal melting values for all 136 possible pairs within the selected 16-peptide interactome. All peptides are distinct from those in our previous quadruple set.²² For a full description of the software, see the [Supporting Information](#).

lacking. We use a combination of known free energies derived via double-mutant analyses for electrostatic $g-e^{+1}$ interactions⁹ and predominantly hydrophobic $a-a'$ interactions^{10,11} and combine these with general amino acid properties such as helical propensity¹² to predict the T_m of a parallel dimeric CC given only the sequences of the constituent peptides. Optimizing these parameters using our growing training set of experimentally tested CCs has allowed us to refine our software to make more accurate predictions that can then be tested on a much larger data set derived using protein arrays.¹³ These new bioinformatics tools for CC prediction began with the bZIP coiled coil interaction prediction algorithm (bCIPA).^{14,15} bCIPA was derived to estimate the T_m of a given parallel dimeric CC using only the primary sequence and was shown to correctly predict 97% of all strong interactions and 95% of all noninteracting pairs using an independent data set of human bZIP proteins.¹³ This prediction was more accurate than a previously published prediction program¹⁶ and utilized very simple and easily adaptable scoring matrices. Unlike related qualitative algorithms,^{13,16} bCIPA makes a quantitative estimate by predicting a T_m value for an interaction between two component polypeptide chains. The approach is distinct from more recent work by the Keating group,^{1,17–19} which makes predictions via complex computational algorithms using integer linear programming and cluster expansion to generate peptide ligands for defined targets and/or off-targets. Although this software does not consider antiparallel dimers, the group has created bespoke software that does.²⁰ Similarly, the Woolfson group derived the CCBUILDER software to study CCs by generating backbones, building in side chains, and providing atomistic models and a range of metrics on which to test their designs.²¹ In contrast to their software, our software is

web-based and user-accessible and differs in that it searches within large user-defined peptide sets to identify and provide quantitative outputs in the form of a T_m to derive heterospecific CC sets while directing against antiparallel CC alignments.

Building on our work in this area,²² we are constructing and expanding a suite of software to meet the goal of identifying very large sets of orthogonal pairs starting from large peptide libraries (Figure 1). In turn, these results are being used to further refine the accuracy of prediction while facilitating a new biological understanding and an expansion in use by the wider scientific community. These user-friendly tools complement experimental work and will allow for the development of designed CC motifs that are highly specific and have a wide range of potential downstream applications mentioned above. By experimentally testing *in silico* predictions, we demonstrate effectiveness in providing new and expanded heterospecific sets while concomitantly refining the software for the design and creation of customized sets that vary in stability according to the needs of the user.

MATERIALS AND METHODS

Design Rationale. The peptide library contained semi-randomized residues at all four a , e , and g positions within the heptad repeat of the 37mers.²² Options of Glu and Lys were included at all e and g positions. Lys was used because its performance is comparable to that of Arg in terms of helicity and the formation of electrostatic interactions but is easier to incorporate into synthetic peptides. Gln, used in previous libraries and designs,^{14,23} has been omitted because it interacts favorably with both acidic and basic residues and is not therefore expected to confer significant specificity to pairings and would therefore be expected to be selected out during the

screening. At *d* positions, Leu was maintained throughout as these are known to assist in driving the formation of parallel and dimeric CC species. At *a* positions, the residues were semirandomized to Asn and Ile. These residues provide the greatest specificity distinction between core position residues based on double-mutant analyses,¹¹ with Asn-Asn (−2.4 kcal/mol) and Ile-Ile pairs (−9.2 kcal/mol) both significantly more favorable than an Asn-Ile pair (−0.5 kcal/mol). These energetic values are anticipated to give a specificity enhancement caused by favorable alignment relative to misaligned residues. Therefore, Asn-Asn pairing confers specificity because the hydrogen bonding benefit outweighs the lack of stability and limits oligomeric states to dimers.^{24,25} Asn-Asn and also Ile-Ile *a-a'* pairs are predicted to stabilize the derived peptides as dimers rather than higher-order oligomers or antiparallel CCs, where Asn-Asn core pairings are also not found.²⁶ This is because *a-a'* and *d-d'* contacts occur in parallel but not antiparallel CCs, meaning that an interaction between equivalent Asn residues in a homodimer will favor a parallel alignment.²⁷ Furthermore, it is anticipated that alignment of Asn residues in core positions will stabilize a particular axial alignment and prevent alternative axial alignments causing unexpected interaction patterns.

In Silico Library Screening. The *in silico* library was created using “Generate Library Sequences” (see the [Supporting Information](#)) to list each user-defined member of the library in a sequential manner. This library was next screened using the “bCIPA Interactome Screen” engine (see the [Supporting Information](#)), which was developed to screen interactomes of sequences using bCIPA^{14,15} and derive a heat map for millions of hypothetical peptide pairs. A 4096-member peptide interactome was reduced to 1536 members by specifying that a minimum of two Asn and two Ile residues is required at *a* positions to assist in imposing specificity. The resulting 1180416 hypothetical pairwise interactions within it were next screened using “Find Pairs” (see the [Supporting Information](#)) to identify groups of four sequences that when placed together would be predicted to form heterospecific dimeric interactions, known as “pairs”. These pairs could then be further screened within the same page (using “Find Quadruples”) to identify groups of eight sequences that when placed together in solution would again be predicted to form four heterospecific dimeric interactions, known as quadruples. Finally, quadruple sets were combined to identify sets of 16 peptides that can form eight heterospecific CCs (“Find Octuples”).

Sequence Screening Protocol. Sequences that met the conditions of the initial constraints described were retained for the interactome screen. These were specificity against homodimerization and the requirement of two Asn residues. The latter has been used previously to create heterospecific sets^{22,28} because it maximizes the potential for specificity in desired pairs where core NI pairings are energetically much less favored than NN or II pairings.¹¹ Elimination of sequences that do not fulfill these requirements at the outset reduces the computational load, allowing even larger libraries to be screened than in the presented example. Each new sequence that satisfied these criteria was added to the array and screened using the “bCIPA Interactome Screen” engine for interaction affinity with every other sequence in the array. This occurred at the time the sequence was added and prevents any repeated calculations, so that each interaction is calculated only once (i.e., not bidirectionally). The results of these calculations were stored in the database, but only if the the affinity of those

interactions exceeded the minimal specified affinity of the desired heterospecific pairs (in this case 70 °C). Thus, this database was a list of pairs of sequences, which could potentially form heterospecific pairings. Interactions in this database, with a T_m greater than the minimum allowed in the input, were paired with each other iteratively, with a computational load-saving requirement that excluded pairs from being screened against one another where those pairs contained any of the same peptides (e.g., an interaction between peptides 1 and 2 could not be paired with an interaction between peptides 1 and 3, because peptide 1 appears in both interactions such that the pairs would not be specific, as there is clear cross talk without needing to quantify the interactions). Potential pairs that did not have any identical sequences were paired iteratively, in a manner similar to identifying the peptide pairs. However, instead of a simple bCIPA calculation, a mini-interactome was created for each potential pair and the T_m calculations of interactions contained therein were checked against a user-specified maximal undesired T_m . Any undesired interactions with a predicted T_m of >20 °C meant that the group of sequences was rejected as a specific pair. In cases in which sequences met these criteria, they were retained as a pair of noninteracting CCs identified in the interactome. Quadruples were next identified by comparing sets of pairs to one another in a similar manner, as done previously (by cross-checking identified noninteracting pairs). However, in the case of quadruples, the increased stringency meant that a higher maximum T_m for an undesired interaction was used, in this case 30 °C, with a minimal ΔT_m (desired minus nondesired) of 40 °C.

Screening Parameters. To generate sets of 16 peptides predicted to form eight heterospecific CCs, the maximal acceptable predicted T_m for homodimers was set to 10 °C (this value dictates the number of nonhomodimeric peptides permitted to proceed into the main screen), the minimal T_m for desired heterodimers to 70 °C, the maximal T_m for undesired heterodimers to 20 °C, and the minimal ΔT_m (desired minus off-target) to 50 °C. Further increasing stringency resulted in fewer initial peptides that progressed to octuples or resulted in many lower-stringency sets [i.e., lower ΔT_m (desired minus off-target)] that therefore took significantly longer to identify. These parameters resulted in the software identifying 42 separate pairs of predicted noninteracting CCs. The highest predicted T_m for desired CCs was 73 °C, and the highest predicted T_m for undesired CCs was 18 °C. Having identified two heterospecific CCs, the program combined pairs to identify 72 sets of four CCs (quadruples). Next, a minimal ΔT_m of 21 °C and a maximal undesired CC T_m of 52 °C were specified within the software. This resulted in the retention of 72 groups of noninteracting quadruples with a lowest desired T_m of 73 °C and a highest undesired CC T_m of 28 °C. Finally, the same parameters were used to combine quadruples in identifying eight CCs (octuples). This resulted in 36 groups of noninteracting quadruples of CCs, with a lowest desired T_m of 73 °C and a highest undesired T_m of 52 °C.

Homodimer Removal. To preserve system resources and to limit the interactome screen to useful search space, sequences that were not expected to produce specific CCs were removed. Search constraints for the interactome excluded all sequences that were predicted to have a homodimeric T_m of >10 °C at the earliest opportunity (as sequences are imported into the script). Sequences retained at this stage were stored in

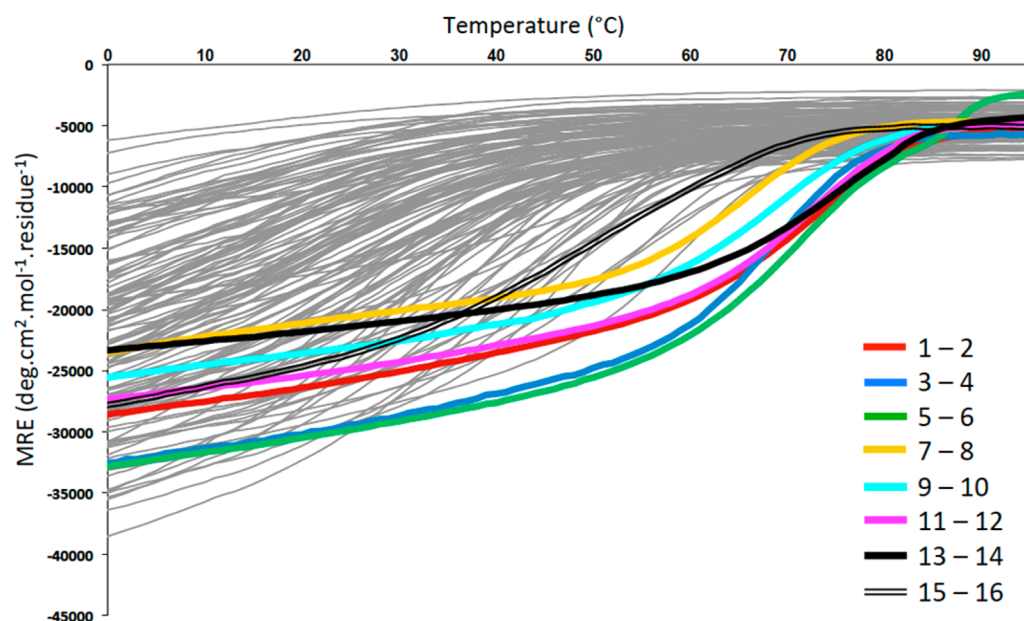


Figure 2. Thermal stability of peptide pairs measured by using the temperature dependence of the CD signal at 222 nm. All 136 peptide pairs are shown, with heterospecific pairs color-coded according to the key. The data demonstrate that with the exception of residues 15 and 16, all desired peptides display T_m values that are higher than that of any measured off-target pair.

a MySQL database, together with the Williams helicity score¹² (to save recalculation).

Antiparallel CC Removal. We have enabled a new feature that searches for and removes homodimers that generate full electrostatic complementarity in the antiparallel orientation. We previously noted that antiparallel dimers were not predicted to form because Asn-Asn core pairings between *a-a'* residues that make the major energetic contribution to CC specificity in the parallel orientation are unable to do so in the antiparallel orientation.²² Rather, buried polar interactions in antiparallel dimers take place between *a-d'* residues and would therefore not be considered possible in this system.^{29,30} This approach has been used previously to direct against antiparallel dimer formation for heterospecific sets.²⁸ However, we previously speculated that this was not enough to direct against potential antiparallel orientations that result in fully complementary electrostatics (i.e., *e-e'* or *g-g'*).²² Directing against full electrostatic complementarity in the antiparallel orientation therefore provides an additional barrier to removing these otherwise permissible antiparallel pairs. It also reduces the search time of the algorithm by increasing the stringency in the selection of the initial sequences that are processed into interactions and consequently reduces the size of the search required to find pairs and quadruples.

Peptide Synthesis. Rink amide ChemMatrix resin was obtained from PCAS Biomatrix, Inc. (Saint-Jean-sur-Richelieu, QC). Fmoc-L-amino acids and 2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) or benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate (PyBOP) were obtained from AGTC Bioproducts (Hessle, U.K.). All other reagents were of peptide synthesis grade and obtained from Thermo Fisher Scientific (Loughborough, U.K.). Peptides were synthesized on a 0.1 mmol scale on a PCAS ChemMatrix Rink amide resin using a Liberty Blue microwave peptide synthesizer (CEM, Matthews, NC) employing Fmoc solid-phase techniques (for a review, see ref 31) with repeated steps of coupling, deprotection, and washing (4×5

mL of dimethylformamide). Coupling was performed as follows: Fmoc amino acid (5 equiv), HBTU OR PyBOP (4.5 equiv), and diisopropylethylamine (10 equiv) in dimethylformamide (5 mL) for 5 min with 35 W microwave irradiation at 90 °C. Deprotection was performed as follows: 20% piperidine in dimethylformamide for 5 min with 30 W microwave irradiation at 80 °C. Following synthesis, the peptide was acetylated, with acetic anhydride (3 equiv) and diisopropylethylamine (4.5 equiv) in dimethylformamide (2.63 mL) for 20 min, and then cleaved from the resin with concomitant removal of side chain-protecting groups by treatment with a cleavage mixture (10 mL) consisting of TFA (95%), triisopropylsilane (2.5%), and H₂O (2.5%) for 4 h at room temperature. Suspended resin was removed by filtration, and the peptide was precipitated using three rounds of crashing in ice-cold diethyl ether, vortexing, and centrifuging. The pellet was then dissolved in a 1:1 MeCN/H₂O mixture and freeze-dried. Purification was performed by reverse-phase high-performance liquid chromatography (RP-HPLC) using a Phenomenex Jupiter Proteo (C18) reverse-phase column (4 μ m, 90 Å, 10 mm inner diameter \times 250 mm length). Eluents used were as follows: 0.1% TFA in H₂O (A) and 0.1% TFA in MeCN (B). The peptide was eluted by applying a linear gradient (at 3 mL/min) of 5 to 70% B over 40 min. Fractions collected were examined by electrospray mass spectrometry, and those found to contain exclusively the desired product were pooled and lyophilized. Analysis of the purified final product by RP-HPLC indicated a purity of >95%.

Circular Dichroism (CD). CD was recorded using an Applied Photophysics (Leatherhead, U.K.) Chirascan CD apparatus using a 200 μ L sample in a CD cell with a 1 mm path length. Samples contained a 150 μ M total peptide (P_t) concentration at an equimolar concentration for heterodimeric solutions (i.e., 75 μ M per peptide) and suspended in 10 mM potassium phosphate and 100 mM potassium fluoride (pH 7) 1 h prior to analysis. The buffer was chosen to be CD-compatible while being close to physiological pH and salt conditions. The

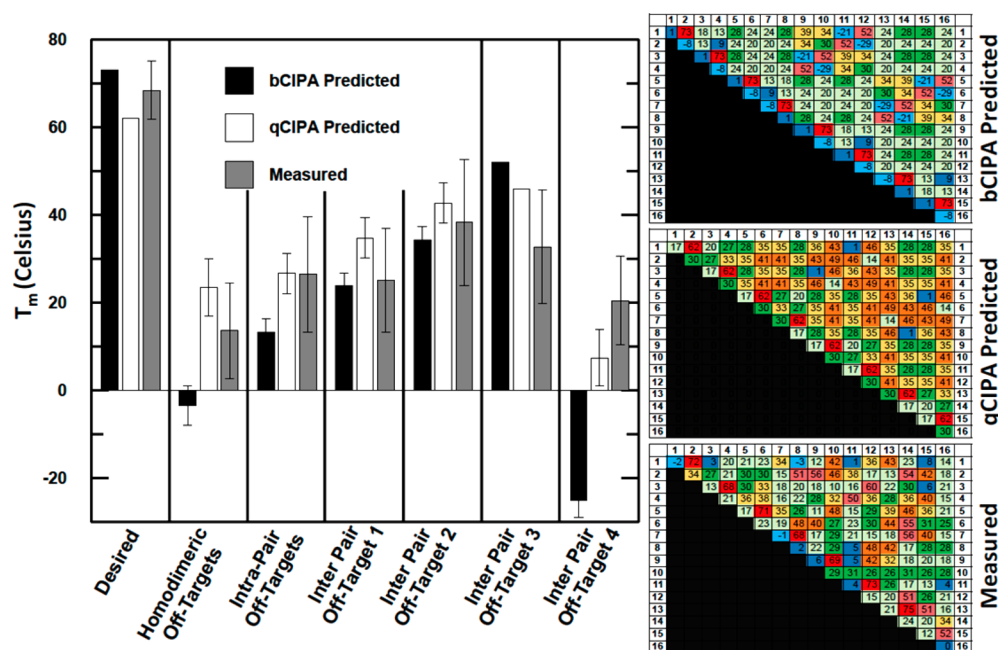


Figure 3. Bar chart and heat maps displaying T_m values predicted by bCIPA and qCIPA as well as those experimentally measured. The values have been grouped according to the core and electrostatic arrangements. These are desired pairs (2II 2NN core, all-attractive electrostatics), homodimeric off-targets (2II 2NN core, all-repulsive electrostatics), intrapair off-targets (4NI core, four attractive and four repulsive electrostatics), interpair off-target 1 (1II 1NN 2NI core, four attractive and four repulsive electrostatics), interpair off-target 2 (2II 2NN core, four attractive and four repulsive electrostatics), interpair off-target 3 (4NI core, all-attractive electrostatics), and interpair off-target 4 (4NI core, all-repulsive electrostatics). qCIPA was derived by using the 22 CCs from our previous work in this area²² as a training set. Instead of core and electrostatic weightings used by bCIPA, qCIPA uses individual weightings for II/NN/IN (core) and EE/KK/EK (electrostatic) arrangements, which results in an improved fit to the training set. See also Tables S1–S7 for a comprehensive list of core and electrostatic combinations that fall into one of the seven categories described above.

CD spectra of samples were scanned between 300 and 190 nm in 1 nm steps, averaging 0.5 s at each wavelength. Three scans at 20 °C were averaged to assess helical levels and CC structure.

Thermal Denaturation Experiments. Thermal denaturations were performed at 150 μ M P_i in 10 mM potassium phosphate and 100 mM potassium fluoride (pH 7) using an Applied Photophysics Chirascan CD instrument. The temperature ramp was set to stepping mode using 1 °C increments and paused for 30 s at each temperature before the ellipticity at 222 nm was measured. For all temperature denaturation experiments, data collection was started at -8 °C, and at this temperature, the peptide solutions remained aqueous. Data collection continued to 95 °C. Data points for thermal denaturation profiles represent the averaged signal after data collection for 4 s. Melting profiles (Figure 2) were $\geq 95\%$ reversible with equilibrium denaturation curves fitted to a two-state model, derived via modification of the Gibbs–Helmholtz equation,^{14,32,33} to yield the melting temperature (T_m). Melting profiles for heterodimers are clearly distinct from averages of constituent homodimeric melts (Figures 2 and 3), indicating that helices form heterodimeric complexes, with the cooperative nature of the melting profiles suggesting an apparent two-state process. T_m values were determined by least-squares fitting of the denaturation assuming a two-state folding model that is widely used for CCs³³ and provided an excellent fit to our data.

Size Exclusion Chromatography. Size exclusion experiments were performed at room temperature using a Superdex Peptide 10/300 GL column (GE Healthcare Life Sciences) by injecting 100 μ L of a 50 or 10 μ M (total peptide concentration) sample in 10 mM potassium phosphate and 100 mM potassium

fluoride (pH 7) at a flow rate of 0.5 mL/min. Elution profiles were recorded via A_{280} .

RESULTS AND DISCUSSION

We previously generated a 1536-member computational library of peptides that were 32 residues in length and successfully screened it using a PHP-based algorithm to predict the formation of four heterospecific parallel dimeric CCs.²² Here we describe screening the resulting 1180416-CC interactome $[(1536 \times 1537)/2]$ using a much faster and more expansive Python-based algorithm that has allowed the identification of many different sets of 16 peptides that when combined are predicted to form eight heterospecific dimeric CCs. When the number of desired heterospecific pairs is doubled from four to eight, the number of off-targets is quadrupled from 32 $[(8 \times 9/2) - 4]$ to 128 $[(16 \times 17/2) - 8]$, making this a particularly challenging task (Figure 1). The algorithm is further improved over the previous version in that it removes peptides predicted to form antiparallel CCs. Dimerization is driven by Lys/Glu options at every *e* and *g* position and Ile/Asn options at every *a* position within the heptad repeat, creating the necessary options to direct the formation of heterospecific CC sets.^{10,11,34} The *d* positions were fixed as Leu to further direct formation of parallel and dimeric CCs,^{24,35} with remaining positions fixed as Ala to promote α -helicity. Screening works by iteratively identifying within a set of sequences which CCs are and are not predicted to form using a set of temperature cutoffs input by the user. The program assigns a predicted T_m for every hypothetical CC within the interactome and creates an associated heat map. Stringency of screening can be directed by inputting the required T_m for desired pairs, as well as the T_m cutoff for homodimeric and heterodimeric off-targets (see

Table 1. Energetic Arrangements Found in Quadruples Are Expanded in Sets of Octuples^a

	quadruples		octuples		octuples with arrangement	predicted T_m range from bCIPA	predicted T_m range from qCIPA	measured T_m range
	core	electrostatics	core	electrostatics				
desired pairs	2II, 2NN $\Delta G = -23.2$	8EK/KE $\Delta G = -9.6$	2II 2NN $\Delta G = -23.2$	8EK/KE $\Delta G = -9.6$	8	73 °C $\Delta G = -32.8$	62 °C	52–75 °C
homodimeric off-targets	2II, 2NN $\Delta\Delta G = 0$	8KK/EE $\Delta\Delta G = +7.2$ to +12.8	2II 2NN $\Delta\Delta G = 0$	8KK/EE $\Delta\Delta G = +7.2$ to +12.8	16	–8 to 1 °C $\Delta\Delta G = +7.2$ to +12.8	17–30 °C	–2 to 34 °C
intrapair off-targets	4 NI $\Delta\Delta G = +21.2$	4KE, 4EE/KK $\Delta\Delta G = +3.6$ to +6.4	4NI $\Delta\Delta G = +21.2$	4KE, 4EE/KK $\Delta\Delta G = +3.6$ to +6.4	16	9–18 °C $\Delta\Delta G = +24.8$ to +27.6	20–33 °C	3–51 °C
interpair off-targets	1II, 1NN, 2NI $\Delta\Delta G = +10.6$	4KE, 4EE/KK $\Delta\Delta G = +3.6$ to +6.4	1II, 1NN, 2NI $\Delta\Delta G = +10.6$	4KE, 4EE/KK $\Delta\Delta G = +3.6$ to +6.4	64	20–28 °C $\Delta\Delta G = +14.2$ to +17	28–41 °C	–3 to 60 °C
			2II 2NN (new) $\Delta\Delta G = 0$	4KE, 4EE/KK (new) $\Delta\Delta G = +3.6$ to +6.4	16	30–39 °C $\Delta\Delta G = +3.6$ to +6.4	36–39 °C	12–60 °C
			4NI (new) $\Delta\Delta G = +21.2$	8KE (new) $\Delta\Delta G = 0$	8	52 °C $\Delta\Delta G = +21.2$	46 °C	10–42 °C
			4NI (new) $\Delta\Delta G = +21.2$	8KK/EE (new) $\Delta\Delta G = +7.2$ to +12.8	8	–29 to –21 °C $\Delta\Delta G = +28.4$ to +34	1–14 °C	1–32 °C

^aThe doubling of the number of desired heterospecific coiled coil pairs leads to a loss of stringency and therefore specificity of interaction as the number of off-targets is quadrupled. The three new interpair off-target combinations are 2II 2NN core with four attractive and four repulsive electrostatic interactions and all-attractive and all-repulsive ES with 4NI core mismatches, resulting in bCIPA estimating the T_m to be 52 or –21/–29 °C, respectively. Shown are the contributions to folding from core and electrostatic interactions, as well as their sums. All free energies are shown in kilocalories per mole and are based on free energy scores derived from a double-mutant analysis.

Materials and Methods and the Supporting Information for more details).

Combining Core and Electrostatic Arrangements To Confer Stability and Specificity. The number of energetic arrangements in the set of octuples (eight CC) is increased from those previously observed (Table 1). Setting the desired CC cutoff T_m as high as possible (70 °C) led the software to arrive at fully complementary electrostatic arrangements and fully optimal 2xII/2xNN core arrangements. For homodimeric off-targets, the same core arrangements were observed as for desired pairs, but with fully repulsive electrostatic ($g\text{-}e^{+1}$) arrangements. For intrapair off-targets (i.e., within a designed interactome of two CCs), the core consisted of a 4NI/IN (fully mismatched) arrangement with four attractive (KE) and four repulsive (EE/KK) electrostatic interactions. However, the interpair off-target arrangements (interactions outside of designed interactomes of two CCs) were expanded in the set of octuples relative to our previously observed combinations.²² In the sets of quadruples (i.e., four CCs; peptides 1–8 or peptides 9–16), we observe a 1II/1NN/2NI core arrangement with the same four attractive (KE) and four repulsive (EE/KK) electrostatic interactions. This was the most commonly observed off-target configuration, accounting for 50% of all off-targets. Via combination of quadruples to arrive at octuples, three additional scenarios were observed (i.e., in CC interactions between peptides from the set of 1–8 and those from the set of 9–16). These additional off-target combinations were (i) an optimized core (2NN 2II) with four attractive and four repulsive electrostatic interactions, leading to a further increase in the predicted stabilities of off-targets, (ii) a fully

mismatched 4NI core arrangement with an all-attractive electrostatic arrangement, and (iii) a fully mismatched 4NI core arrangement with an all-repulsive electrostatic arrangement. The last two scenarios were predicted by bCIPA as being either very stable (52 °C) or very unstable (–29/–21 °C), which we believed to be over- or underestimated, respectively. Therefore, a set of heterospecific octuples could be identified, albeit with an overall stringency lower than for that for quadruples because of the increased stability of off-targets highlighted above (Table 1 and Figure 3). The data from an exhaustive low-complexity set of options (i.e., all Asn/Ile core options and all Glu/Lys electrostatic options) are therefore sufficient to generate the required number of unique core and electrostatic arrangements for the creation of such larger heterospecific sets. Moreover, the large amount of data generated in predicting and experimentally testing an octuple heterospecific set has allowed creation of a more refined version of software. Subsequently, this can be used to predict heterospecific interactomes for this particular subset of CC. This would strengthen prediction of affinity for individual pairs and therefore allow the creation of both larger and more accurate heterospecific sets.

Experimental Characterization of Coiled Coils. To demonstrate that *in silico*-generated sequences are specific *in vitro*, 16 peptides predicted to form eight heterospecific CCs were synthesized and characterized. CD spectra and associated thermal denaturation experiments were used to establish that all samples displayed characteristic α -helical profiles and to determine the T_m value for each CC within the 16-peptide interactome, and therefore the relationship between predicted

and measured values (Figure 3). In these experiments, seven of eight CCs predicted to be heterospecific were verified experimentally, with predicted T_m values of 73 °C found to be accurate within 5 °C, with the exception of that of CC 15–16, which was 21 °C lower than expected. The 128 off-target interactions had predicted T_m values of –8 to 52 °C, which were measured to range from –3 to 60 °C, ensuring that the 1–2, 3–4, 5–6, 7–8, 9–10, 11–12, and 13–14 interactome was heterospecific as designed, with all 98 off-targets disfavored. As one can see, differences in molar ellipticity at premelt temperatures reflect the fact that helicity is only one determinant of CC stability; side-chain preferences at core and electrostatic positions are other key determinants that bCIPA uses when determining the T_m .

To further demonstrate correct peptide pairing, size exclusion chromatography (SEC) was used (Table S8). As shown previously,²² monomeric elution profiles are superimposable and occur at 20 min for all homomeric solutions, indicating that all 16 peptides are monomeric at room temperature and a total peptide concentration of 50 μ M. In contrast, at 50 μ M, the profiles for all eight desired heterodimers eluted at 19 min. In both cases, the elution profiles were consistent with predicted monomer/dimer patterns. In addition, a number of off-target samples were run, demonstrating that those with a measured T_m of <50 °C were monomeric at 50 μ M whereas those with a T_m of >50 °C ran as a dimer at a total peptide concentration of 50 μ M. However, at a concentration of 10 μ M, all desired peptides (T_m = 68–75 °C), with the exception of 15–16 (T_m = 52 °C), were found to remain in complex as dimers. At the same concentration, all nine off-target samples with a measured T_m of >50 °C (i.e., T_m = 51–60 °C) were found to have shifted to monomeric samples. As SEC controls, two peptides of similar length that have been previously characterized and shown to exist either in monomer form (20 min) or as a parallel dimeric CC (19 min) were selected.^{14,22}

Comparison of Predicted and Observed Data (bCIPA).

Throughout the process, T_m values for predicted and observed pairs were compared to improve the accuracy of prediction. Upon comparison of CCs selected together in pairs, where interactomes consist of 10 potential CCs (e.g., peptide sets 1–4, 5–8, 9–12, or 13–16), there was generally an excellent correlation. This was the case in 22 previously characterized CCs (overall r^2 = 0.6; ²² r^2 = 0.69 for peptide set 1–4 within) as well as when the fit was applied to newly derived peptides sets 1–4, 5–8, 9–12, and 13–16 within the 136 CCs. The strong observed correlations (r^2 = 0.70–0.82) relate to the fact that desired pairs have very high T_m values (e.g., two CCs of approximately 73 °C), and four homodimeric off-targets have very low T_m values (four potential CCs of approximately 0 °C). The four remaining members of each *in silico*-selected four-peptide interactome display predicted T_m values in the range of 20–30 °C. Therefore, each resulting 10-CC interactome contains a wide T_m range to which the subsequent fit is strong. Fitting to the off-targets more generally (particularly interpair off-targets) is more challenging because the predicted ΔT_m is very narrow (e.g., typically just 10 °C for the majority of interpair off-targets). This means that similar variations in the predicted temperatures yield much lower r^2 values. However, importantly, the predicted versus observed T_m values for each type of interaction demonstrate that the general goal of heterospecificity is achieved (Figure 3).

Issues with the Computational Approach. Despite successfully demonstrating that PHP-based bCIPA software can predict many hypothetical sets of octuples, we experienced several issues that limited its further implementation. One was speed; it was time-consuming for larger peptide sets and not expansive enough to identify larger numbers of these sets by relaxing the screening parameters. In addition, antiparallel options were not removed. Finally, the accuracy of the bCIPA prediction algorithm left room for improvement in these low-sequence diversity peptides. These problems have been addressed by moving away from PHP-based software to Python-based architecture. Previously, results were uploaded and processed on an external Web server, causing the program to become slow as the number of peptides increased, limiting the number that could be progressed within the search. Using a Python-based system, outputs are now simply written as text files and saved locally by the user, saving computational resources in the process. For instance, increasing the permitted homodimeric T_m by only a few degrees Celsius substantially increases the size of the interactome to be searched. The Python-based approach has allowed these larger data sets to be processed in much less time. The initial check now removes potential antiparallel homodimers by removing those that result in an all-attractive electrostatic component. This additionally prevents potential heterodimeric antiparallel CCs from entering the interactome search. Removing potential parallel homodimers and antiparallel homo- and heterodimers initially restricts the number of peptides that enter the main interactome search, further reducing the redundancy of the system.

Refining bCIPA To Improve T_m Prediction for Specific Residue Pairings (qCIPA). We have previously shown that bCIPA can accurately predict the thermal stability of CC pairs that are diverse in sequence^{14,15,14} and that it can be used to generate *in silico* interactome predictions to guide the derivation of heterospecific CC sets.²² The utility of bCIPA was demonstrated using a small eight-peptide interactome to derive four parallel dimeric CCs that were heterospecific when combined, despite 32 off-target CCs that could potentially associate. Using a completely new set of peptides, we expanded this approach to a 16-peptide interactome. In doubling the number of desired heterospecific CCs from 4 to 8, the number of off-targets quadruple from 32 to 128, leading to a significant increase in the complexity of the design process. In turn, as the number of attractive and repulsive permutations becomes exhausted, higher-stability off-targets must be included in the interactome (see above), leading to a decrease in the stringency of the system. Thus, predicting larger heterospecific sets is a challenging task that requires a high accuracy of prediction because any decrease in stringency will increase the likelihood of identifying off-targets with stabilities similar to those of the desired pairs. We have been largely successful in these aims and have arrived at an interactome of 14 peptides that form seven heterospecific CCs despite 98 potential CC off-targets. In particular, CC 15–16 was found to display a lower T_m that is close to those of some of the off-target interactions. The results from this study further highlight the strengths and weaknesses of bCIPA. We have used our previous interactome data set of eight independent peptides as well the 16 newly predicted peptides presented here to facilitate the creation of new customized software, known as qCIPA. qCIPA predicts the T_m with accuracy greater than that of bCIPA for the subset of CCs we describe here that are closely related in sequence (Figure 3).

qCIPA was devised using 22 previously characterized CCs as a training set ($r^2 = 0.69$). The resulting fit was then applied to the 136-CC test set resulting from the 16-peptide interactome described here, with no sequence repetition between the two sets. The resulting correlation coefficients therefore provide a direct comparison with bCIPA (Table S9). On average, qCIPA provides a 3 °C improvement in prediction using the 136-CC test set. In reducing the test set interactome to its substituent interactions, we find a 9 °C improvement in predicting the 16 homodimers, a 4 °C improvement in predicting the 16 intrapair off-targets, and a 3 °C improvement in predicting the 96 interpair off-targets. This comes at a cost of a 3 °C deterioration in predicting the heterospecific pair interactions (Figure 3). Considering the end goal of creating heterospecific sets, in which off-target interactions should be carefully avoided, this overall increase in accuracy is a welcome step forward. qCIPA works in a manner similar to that of bCIPA in that the T_m is calculated as a function of core pairings (C), electrostatic pairings (ES), and helical propensity (HP). For bCIPA

$$T_m = a \times \sum \text{HP} + bC + c \times \text{ES} + d \quad (1)$$

where $a = 81.33$, $b = -10.18$, $c = -4.78$, and $d = -29.13$.¹⁵ bCIPA considers a wide range of residues at positions a , d , e , and g (LINVRKT, LINVRKT, KRDEQNALT, and KRDEQNALT, respectively). In contrast for qCIPA, the options at these positions are limited to IN, L, EK, and EK. This allows the six exact pairings at a - a' and g - e'^{+1} to be explicitly described. Therefore, in the equation used to determine T_m , the core and electrostatic components are expanded so that each interaction has its own coefficient. For qCIPA

$$T_m = a \times \sum \text{HP} + b \times \text{II} + c \times \text{IN} + d \times \text{NN} + e \times \text{EE} + f \times \text{EK} + g \times \text{KK} + h \quad (2)$$

where $a = 4.16$, $b = -1.75$, $c = 11.78$, $d = -5.24$, $e = -11.30$, $f = -0.97$, $g = -76.22$, and $h = 30.18$. As observed in Figure 3, these changes result in an improved fit to the 22-CC training set (i.e., the best fit that is used to derive values of a - h ; overall $r^2 = 0.69$). Having obtained the values listed above, we then obtained the consequent fits to the CC pairs in the 136-CC test set (r^2 values of 0.89, 0.69, and 0.84 for pairs 1-4, 5-8, and 9-12, respectively, and 0.59 for set 13-16).

As the data set continues to grow, we predict that it will be possible to take further parameters into account, such as sequence specific context where the core and electrostatic contributions are equivalent but their positioning within each helix leads to increased or decreased stability above or below what is otherwise predicted. We have seen this previously for positive or negative residues at the helix termini that serve to stabilize or destabilize the helix macrodipole leading to over- or underestimated stability.^{22,36,37} At present, there are insufficient data to build these predictions into our models, although general patterns within the data are emerging (see below).

Comparing Old and New Approaches. To generate octuples (including the set presented here), the PHP-based bCIPA software was used to generate two heterospecific CCs and consequently quadruples that were predicted to remain heterospecific when all component peptides are combined. This resulted in 72 unique sets of two CCs (setting the minimal delta to 50 °C and maximal off-targets permitted to progress into the interactome search to 20 °C) and 144 sets of quadruples (setting the minimal delta to 40 °C and maximal off-targets to 30 °C). By repeating these steps while removing

potential antiparallel pairs, we found the numbers decreased to 42 unique sets of two CCs and 72 unique sets of quadruples. To continue using this approach, unique sets of quadruples were screened against each other to identify unique sets of octuples. However, this was time-consuming and led to only 36 hypothetical octuple sets.

Using the faster Python-based qCIPA software now proposed, the algorithm created heterospecific sets by screening one CC against another, one CC at a time [e.g., one CC → two → three → four (quadruples) → five → six → seven → eight (octuples)], until no further unique sets can be identified. For example, the 1536 peptides scanned for heterospecific pairs took 6 s using qCIPA in Python. In contrast, bCIPA benchmarked at 42 s on the same machine when using PHP. This has led to many more unique sets being identified because the stringency in taking smaller increments is much lower and therefore more peptides are permitted to progress at each step. For example, using slightly less stringent parameters [the maximal T_m for undesired heterodimers and ΔT_m (desired minus off-target) (see Table 2)], the 510 unique sets of pairs

Table 2. Numbers of Unique Sets Identified Using the Previous PHP-Based Approach versus a Python-Based Approach^a

tuple	no. of sets in the PHP system (parameters used)	no. of sets in the Python system (parameters used)	no. of sets in the Python system (parameters used)
1	36 (70/10/20/50)	36 (70/10/52/21)	36 (66/14/50/23)
2	42	510	492
3	N/A	3708	3264
4	72 (70/10/52/21)	15171	11067
5	N/A	36204	18180
6	N/A	51450	11430
7	N/A	45456	0
8	36	27501	0
9	N/A	11904	0
10	N/A	3552	0
11	N/A	648	0
12	0	54	0

^aBoth data sets implement the exclusion of predicted high-affinity antiparallel dimers and specify a minimum and maximum of two Asn residues to confer maximal core specificity upon selected sets. For PHP, the approach of combining sets (e.g., 2 → 4 → 8) was necessary to reduce the computing time required to generate results. In contrast, Python adds one CC at a time, resulting in much larger numbers of predicted heterospecific sets. By using the settings required at the quadruple stage from the beginning (column 2 vs column 1), we found that once again more CCs were permitted to progress through each round to arrive at duodecuples, further demonstrating the benefit of early redundancy in the system. Further increasing the stringency (minimal ΔT_m) resulted in no heterospecific sets beyond sextuples (column 3). Column 2 took Python approximately 4 days to run on a single personal computer. Shown in parentheses are the minimal desired T_m , the minimal homodimer T_m , the minimal off-target T_m , and the minimal ΔT_m (from left to right, respectively), requested for each round.

created 15171 unique sets of quadruples. This 210-fold increase in the number of unique sets created 27501 unique sets of octuples, >760-fold more than previously identified. Coupled with this iterative approach, relaxing the homodimer stringency from the very start of the search procedure (Table 2) can be continued. Although not tested experimentally, we have taken this iterative process as far as 54 unique sets of predicted

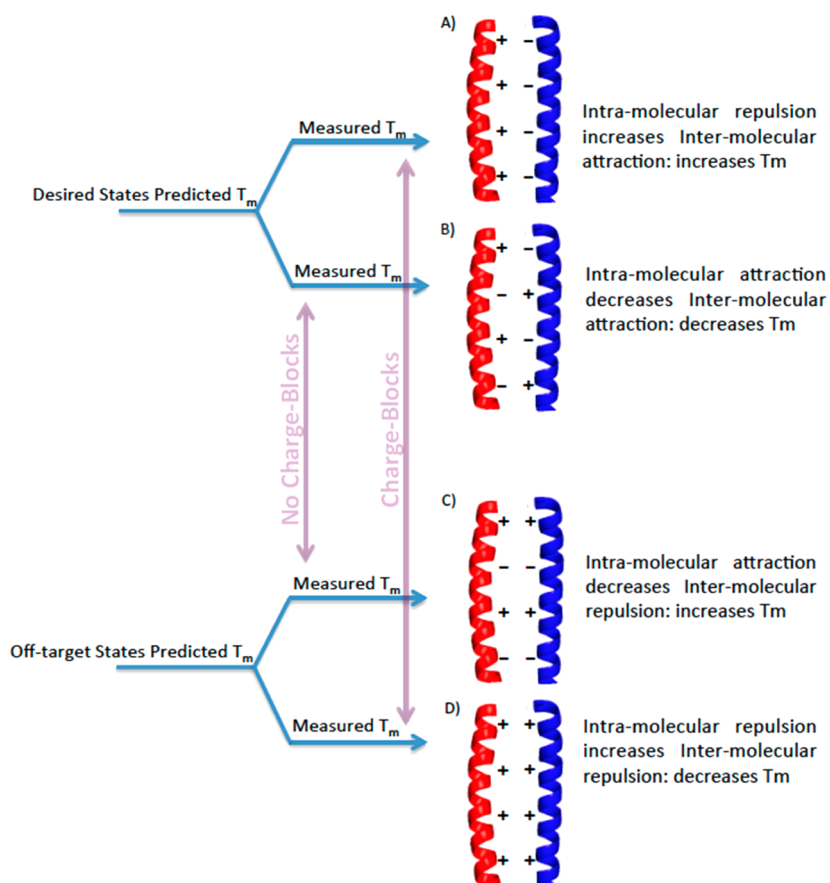


Figure 4. Effect of electrostatic charge blocks on the predicted T_m . For desired pairs, blocks of three or four consecutive same-charge residues at e or g positions on each peptide led to increased stability versus that purely based on the sum of the core and electrostatic components. We speculate that these “charge blocks” have two benefits. (i) They increase stability for these desired pairs by promoting intermolecular attraction (e.g., A vs B). (ii) In a similar but opposite manner, charge blocks decrease the stability of off-target CCs by promoting intermolecular repulsion (e.g., C vs D). The net effect is therefore that $\Delta T_m(\text{desired} - \text{off-targets})$ is increased when “charge blocks” are introduced (i.e., A-D > B-C).

duodecuples (12's), where none could be identified using our previous PHP approach.²² It is important to note that while every set is unique, there are many instances of the same peptide occurring within multiple sets. This apparent redundancy in the search procedure is, however, necessary to ensure that sequences are retained during each iteration and that the highest possible number of heterospecific CC sets can be identified going forward. Relaxing the stringency further will increase this number until all core and electrostatic arrangements have been saturated, while significantly increasing the search time from several days (e.g., 4 days in the case described above) to many weeks on a standard personal computer.

Sequence Specific Context. *Core.* If we ignore sequence context, then many pairs appear to be energetically identical in terms of core and electrostatic contributions (using helix propensity¹² and core^{10,11} and electrostatic⁹ scores calculated by Vinson and co-workers). The absence of sequence context calculations is reflected in the lack of diversity in predicted T_m values. While we found it is currently difficult to build into a qCIPA feature explicitly, we observe some general rules relating to sequence context that can be taken into account in future rounds of design. This could be achieved, for example, by allowing the software to search libraries that conform to these rules in the first instance, such that the sequence specific peptides no longer need to be explicitly “searched for”. In analyzing the data, by grouping sequences with identical electrostatic arrangements, we are able to make some limited

interpretations regarding the effect of core arrangements. (1) An NN II NN II arrangement appears to lead to a stability that is greater than that of an II NN II NN arrangement. For example O5–6 > O15–16 ($\Delta T_m = 19$), O13–14 > O7–8 ($\Delta T_m = 7$), O5–5 > O15–15 ($\Delta T_m = 5$), O14–14 > O8–8 ($\Delta T_m = 22$), O6–6 > O16–16 ($\Delta T_m = 23$), and O13–13 > O7–7 ($\Delta T_m = 22$). (2) In contrast to this, inspection of the data suggests that NN NN II II and II II NN NN are energetically equivalent. For example, O9–10 ~ O3–4 and O1–2 ~ O11–12. However, both of these arrangements are predicted to stabilize desired states and off-targets by an equal amount, meaning that there is no preferential core arrangement in maximizing ΔT_m values and therefore in achieving heterospecific CCs.

Electrostatics. As there are many examples of alternative electrostatic arrangements with identical core arrangements (Tables S1–S7), we are able to make some general observations. Upon normalization for identical cores, the following findings become apparent. (1) For desired pairs (Table S1), blocks with the same charge on residue e or g of each peptide led to a stability increased versus that purely based on the sum of the core and electrostatic components.^{9–11} For example, O13–14 > O5–6 ($\Delta T_m = 4$) and O7–8 > O15–16 ($\Delta T_m = 16$), which together suggest that the electrostatic $g-e'^{+1}$ KE KE KE EK arrangement is more stable than the EK EK KE EK arrangement. Similarly, O1–2 > O9–10 ($\Delta T_m = 3$) and O11–12 > O3–4 ($\Delta T_m = 5$), both suggesting that the KE EK

EK EK arrangement is more stable than the EK KE EK EK arrangement. Taking this further, we found both O1–2 and O9–10 are favored over Q1–2 ($\Delta T_m = 9$ and 6, respectively), suggesting that the EK KE KE EK arrangement is less stable than the KE EK EK EK or EK KE EK EK arrangement. Collectively, this suggests that for desired pairs, intramolecular repulsion between heptads strengthens intermolecular attraction and therefore increases CC stability. We speculate that these intramolecular “charge blocks” at positions *e* and *g* within component helices increase the stability for these desired pairs by promoting intermolecular attraction. This may be due to the fact that intramolecular repulsion between *e-e*⁺ or *g-g*⁺ residues helps to strengthen intermolecular attraction between *g-e*⁺ pairs. (2) Similarly, for intrapair off-targets (Table S3), placing opposing intermolecular charge repulsions next to each other (i.e., E followed by K at consecutive *e* or *g* positions within the same peptide) is more stabilizing than same polarity charge repulsions [e.g., O5–7 or O13–15 > O2–4, O10–12, O1–3, or O9–11 ($\Delta T_m = 9$ –48); O2–3 or O9–12 > O6–7, O13–16, O5–8, or O14–15 ($\Delta T_m = 1$ –26)]. This suggests that for off-targets, intramolecular attraction weakens intermolecular repulsion and increases CC stability. Similarly, when there is both intra- and intermolecular repulsion (i.e., ++ or --), the T_m is decreased. We speculate that alternating charges at *e* or *g* positions promotes intramolecular attraction and leads to decreased *g-e*⁺ intermolecular repulsion. These effects increase the stability for these intrapair off-targets. This pattern is observed throughout the off-target sets. Alternating intramolecular charges are therefore to be disfavored for both desired states and off-targets when designing heterospecific sets and function in an opposite but analogous way (Figure 4). (3) For homodimeric off-targets (Table S2), negative intermolecular charge repulsions toward the N-terminus (i.e., E-E *g-e*⁺ pairs) and positive charge repulsions (i.e., K-K *g-e*⁺ pairs) at the C-terminus generally generated increased stability. Reversing this pattern generally destabilizes the CCs. In general, it is more stabilizing for the CC to have a negative N-terminus than to have a positive C-terminus [Q1–1 > Q2–2 ($\Delta T_m = 12$), O9–9 > O1–1 ($\Delta T_m = 8$), O3–3 > O11–11 ($\Delta T_m = 9$), O2–2 > O10–10 ($\Delta T_m = 5$), and Q8–8 > Q7–7 ($\Delta T_m = 16$)].³⁷ The effect of placing a positive repulsive pair at the C-terminus is not clear. This pattern of negative charge at the N-terminus and positive charge at the C-terminus adding stability generally holds for interfamilial off-targets (e.g., O12–14 vs O3–5 or O4–6) and is most pronounced when the electrostatics are fully repulsive (e.g., O5–15 vs O6–16).

On the basis of these findings, in the future solid blocks of three or more E/K residues (i.e., at three consecutive *e* or *g* positions) should be included in peptide library designs because these blocks will assist in stabilizing desired pairs while concomitantly destabilizing off-targets, leading to a favorable increase in ΔT_m (desired minus off-target) (Figure 4). In addition, introducing E at the N-terminus and (less so) K at the C-terminus will further aid stability. Alternating charge repulsions on the same helix should be avoided because they promote intrahelical electrostatics; this will have the effect of both weakening intermolecular repulsion for off-targets while also weakening the intermolecular electrostatic attractions in the desired states. Although these observations present general trends, it is difficult to predict the magnitude of the effects in building sequence specific context into stability prediction models. Nonetheless, creating libraries that conform to these “charge block” rules in the first instance means they no longer

need to be explicitly searched for. Rather, by defining permitted *e* and *g* charge block arrangements (i.e., EEEE/KKKK/EEEE/EKKK/EEEE/KKKE) with the same core arrangement as previously specified [i.e., $(6 \times 7/2) \times 6$ cores = 126-member library], we are able to screen an interactome of 8001 potential CCs. This resulted in the identification of 12 sets of decuples (10 CCs) based on the same cutoff parameters that were used in Table 2.

Improvements in the speed and flexibility of the software mean that many new avenues of *in silico* screening are now possible, with key patterns in the interaction profiles visible from an observational level (Table 1). The added role of sequence context is of interest as it can further improve the prediction of heterospecific peptides by being added to an increased energy gap between desired and nondesired CCs. To further analyze and predict interaction stability based on these patterns, a larger training set would be required.

CONCLUSION

We have increased the capacity of our predictive algorithm to identify a set of 16 new peptides that can form eight heterospecific CC pairs. Of these, seven have been demonstrated to function as predicted. To the best of our knowledge, this is the largest heterospecific set of designed peptides created to date. In expanding the predicted heterospecific set from eight peptides²² to 16, we have accomplished the following. (1) By necessity, we have increased both the speed and the utility of the algorithm. Although we have stopped at octuples, by continuing with our Python-based approach, we have expanded our predicted heterospecific CC setup to duodecuples (12 CCs) using the current library. (2) We have implemented the removal of peptides predicted to form antiparallel CCs (i.e., those that can adopt fully complementary *e-e*⁺ or *g-g*⁺ electrostatic pairs by the algorithm). (3) We have robustly demonstrated both the need for such software and its utility in directing against the expanding number of lower-energy off-targets, in this case from 32 to 128. The new Python-based algorithm predicts that we can further expand the number of peptides to at least 24. This would generate 12 heterospecific CCs, with 288 CC off-targets, using the same two core (Ile/Asn) and electrostatic (Glu/Lys) residue options. (4) We have used our data set of >170 CCs to identify electrostatic “charge blocks” (Figure 4). These aid the *de novo* design of specificity by serving to increase the stability of desired pairs, while concomitantly decreasing off-target stability. Designing charge blocks in future CC-based systems will assist in ensuring that designed peptide sets achieve their desired heterospecificity. Incorporating these and other emerging sequence context-based rules for otherwise energetically equivalent CCs into prediction models will further ensure that off-targets are disfavored while increasing the predicted stability of desired pairs. (5) Lastly, we believe that the heterospecific peptide sequences generated and the tools used to identify them will also be of use to the synthetic biology community. As more data become available, we will expand the size of both training and test sets to further increase CC prediction accuracy. The software and peptides derived from the study, as well as the approach more widely, have the potential to be applied in a variety of downstream applications that include hydrogels, increased complexity nanocages, PPI inhibitors, and peptide tags for uses as molecular probes.^{3–5}

Our aim to derive a heterospecific interactome using 16 peptides was partially achieved; seven of the eight CCs are

shown to be heterospecific. However, observations from the expanded data set have given rise to a significant increase in the accuracy of CC prediction. Incorporating emerging rules into qCIPA selection to screen and select large heterospecific peptide sets represents a significant advance toward designing interactomes that are more likely to be exquisitely specific. In the future, it may be possible to further improve the accuracy of specificity prediction by taking into account additional coupling energies or by accounting for the context dependence of additional residue pair interactions.^{38,39} We believe that these findings make important contributions to the question of how primary sequence governs the stability and specificity of quaternary structures and in the derivation of peptide building blocks to modulate PPIs as well as tools for the synthetic biology community.

■ ASSOCIATED CONTENT

■ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.7b00047.

Details about the software (including web links), peptide sequences, and additional information about the parameters used to identify octuples (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: j.mason@bath.ac.uk.

ORCID

Jody M. Mason: 0000-0002-4118-1958

Author Contributions

J.M.M. suggested and supervised the work. J.M.M. and R.O.C. designed and synthesized the peptides. R.O.C. acquired the CD data. R.O.C. and J.M.M. analyzed the CD data. R.O.C. and A.L. wrote and implemented the software. A.S.P. acquired and analyzed the SEC data. J.M.M. wrote the paper with input from all authors.

Funding

J.M.M. is grateful to Cancer Research UK for a Career Establishment Award (C29788/A11738) and to the Engineering and Physical Sciences Research Council for an Overseas Travel Grant (EP/M001873/2).

Notes

The authors declare no competing financial interest.

■ REFERENCES

- (1) Reinke, A. W., Grant, R. A., and Keating, A. E. (2010) A synthetic coiled-coil interactome provides heterospecific modules for molecular engineering. *J. Am. Chem. Soc.* 132, 6025–6031.
- (2) Boyle, A. L., Bromley, E. H., Bartlett, G. J., Sessions, R. B., Sharp, T. H., Williams, C. L., Curmi, P. M., Forde, N. R., Linke, H., and Woolfson, D. N. (2012) Squaring the circle in peptide assembly: from fibers to discrete nanostructures by de novo design. *J. Am. Chem. Soc.* 134, 15457–15467.
- (3) Shlizerman, C., Atanassov, A., Berkovich, I., Ashkenasy, G., and Ashkenasy, N. (2010) De Novo Designed Coiled-Coil Proteins with Variable Conformations as Components of Molecular Electronic Devices. *J. Am. Chem. Soc.* 132, 5070–5076.
- (4) Gradisar, H., Bozic, S., Doles, T., Vengust, D., Hafner-Bratkovic, I., Mertelj, A., Webb, B., Sali, A., Klavzar, S., and Jerala, R. (2013) Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments. *Nat. Chem. Biol.* 9, 362–366.
- (5) Fairman, R., and Akerfeldt, K. S. (2005) Peptides as novel smart materials. *Curr. Opin. Struct. Biol.* 15, 453–463.
- (6) Dill, K. A., Ozkan, S. B., Weikl, T. R., Chodera, J. D., and Voelz, V. A. (2007) The protein folding problem: when will it be solved? *Curr. Opin. Struct. Biol.* 17, 342–346.
- (7) Mason, J. M., and Arndt, K. M. (2004) Coiled coil domains: stability, specificity, and biological implications. *ChemBioChem* 5, 170–176.
- (8) Woolfson, D. N. (2005) The design of coiled-coil structures and assemblies. *Adv. Protein Chem.* 70, 79–112.
- (9) Krylov, D., Barchi, J., and Vinson, C. (1998) Inter-helical interactions in the leucine zipper coiled coil dimer: pH and salt dependence of coupling energy between charged amino acids. *J. Mol. Biol.* 279, 959–972.
- (10) Acharya, A., Ruvinov, S. B., Gal, J., Moll, J. R., and Vinson, C. (2002) A heterodimerizing leucine zipper coiled coil system for examining the specificity of a position interactions: amino acids I, V, L, N, A, and K. *Biochemistry* 41, 14122–14131.
- (11) Acharya, A., Rishi, V., and Vinson, C. (2006) Stability of 100 homo and heterotypic coiled-coil a-a' pairs for ten amino acids (A, L, I, V, N, K, S, T, E, and R). *Biochemistry* 45, 11324–11332.
- (12) Williams, R. W., Chang, A., Juretic, D., and Loughran, S. (1987) Secondary structure predictions and medium range interactions. *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* 916, 200–204.
- (13) Newman, J. R., and Keating, A. E. (2003) Comprehensive identification of human bZIP interactions with coiled-coil arrays. *Science* 300, 2097–2101.
- (14) Mason, J. M., Schmitz, M. A., Muller, K. M., and Arndt, K. M. (2006) Semirational design of Jun-Fos coiled coils with increased affinity: Universal implications for leucine zipper prediction and design. *Proc. Natl. Acad. Sci. U. S. A.* 103, 8989–8994.
- (15) Hagemann, U. B., Mason, J. M., Muller, K. M., and Arndt, K. M. (2008) Selectional and mutational scope of peptides sequestering the Jun-Fos coiled-coil domain. *J. Mol. Biol.* 381, 73–88.
- (16) Fong, J. H., Keating, A. E., and Singh, M. (2004) Predicting specificity in bZIP coiled-coil protein interactions. *Genome Biol.* 5, R11.
- (17) Thompson, K. E., Bashor, C. J., Lim, W. A., and Keating, A. E. (2012) SYNZIP protein interaction toolbox: in vitro and in vivo specifications of heterospecific coiled-coil interaction domains. *ACS Synth. Biol.* 1, 118–129.
- (18) Grigoryan, G., Reinke, A. W., and Keating, A. E. (2009) Design of protein-interaction specificity gives selective bZIP-binding peptides. *Nature* 458, 859–864.
- (19) Potapov, V., Kaplan, J. B., and Keating, A. E. (2015) Data-driven prediction and design of bZIP coiled-coil interactions. *PLoS Comput. Biol.* 11, e1004046.
- (20) Negron, C., and Keating, A. E. (2014) A Set of Computationally Designed Orthogonal Antiparallel Homodimers that Expands the Synthetic Coiled-Coil Toolkit. *J. Am. Chem. Soc.* 136, 16544–16556.
- (21) Wood, C. W., Bruning, M., Ibarra, A. A., Bartlett, G. J., Thomson, A. R., Sessions, R. B., Brady, R. L., and Woolfson, D. N. (2014) CCBUILDER: an interactive web-based tool for building, designing and assessing coiled-coil protein assemblies. *Bioinformatics* 30, 3029–3035.
- (22) Crooks, R. O., Baxter, D., Panek, A. S., Lubben, A. T., and Mason, J. M. (2016) Deriving Heterospecific Self-Assembling Protein-Protein Interactions Using a Computational Interactome Screen. *J. Mol. Biol.* 428, 385–398.
- (23) Crooks, R. O., Rao, T., and Mason, J. M. (2011) Truncation, Randomization, and Selection: Generation of a Reduced Length c-Jun Antagonist That Retains High Interaction Stability. *J. Biol. Chem.* 286, 29470–29479.
- (24) Harbury, P. B., Zhang, T., Kim, P. S., and Alber, T. (1993) A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science* 262, 1401–1407.
- (25) Oakley, M. G., and Kim, P. S. (1998) A buried polar interaction can direct the relative orientation of helices in a coiled coil. *Biochemistry* 37, 12603–12610.

- (26) Gurnon, D. G., Whitaker, J. A., and Oakley, M. G. (2003) Design and characterization of a homodimeric antiparallel coiled coil. *J. Am. Chem. Soc.* 125, 7518–7519.
- (27) O'Shea, E. K., Klemm, J. D., Kim, P. S., and Alber, T. (1991) X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science* 254, 539–544.
- (28) Gradisar, H., and Jerala, R. (2011) De novo design of orthogonal peptide pairs forming parallel coiled-coil heterodimers. *J. Pept. Sci.* 17, 100–106.
- (29) McClain, D. L., Woods, H. L., and Oakley, M. G. (2001) Design and characterization of a heterodimeric coiled coil that forms exclusively with an antiparallel relative helix orientation. *J. Am. Chem. Soc.* 123, 3151–3152.
- (30) Oakley, M. G., and Hollenbeck, J. J. (2001) The design of antiparallel coiled coils. *Curr. Opin. Struct. Biol.* 11, 450–457.
- (31) Fields, G. B., and Noble, R. L. (1990) Solid-Phase Peptide-Synthesis Utilizing 9-Fluorenylmethoxycarbonyl Amino-Acids. *Int. J. Pept. Protein Res.* 35, 161–214.
- (32) Elwell, M. L., and Schellman, J. A. (1977) Stability of phage T4 lysozymes. I. Native properties and thermal stability of wild type and two mutant lysozymes. *Biochim. Biophys. Acta, Protein Struct.* 494, 367–383.
- (33) Mason, J. M., Hagemann, U. B., and Arndt, K. M. (2007) Improved stability of the Jun-Fos Activator Protein-1 coiled coil motif: A stopped-flow circular dichroism kinetic analysis. *J. Biol. Chem.* 282, 23015–23024.
- (34) O'Shea, E. K., Lumb, K. J., and Kim, P. S. (1993) Peptide 'Velcro': design of a heterodimeric coiled coil. *Curr. Biol.* 3, 658–667.
- (35) Woolfson, D. N., and Alber, T. (1995) Predicting oligomerization states of coiled coils. *Protein Sci.* 4, 1596–1607.
- (36) Cochran, D. A. E., and Doig, A. J. (2001) Effect of the N2 residue on the stability of the alpha-helix for all 20 amino acids. *Protein Sci.* 10, 1305–1311.
- (37) Doig, A. J., and Baldwin, R. L. (1995) N- and C-capping preferences for all 20 amino acids in alpha-helical peptides. *Protein Sci.* 4, 1325–1336.
- (38) Steinkruger, J. D., Woolfson, D. N., and Gellman, S. H. (2010) Side-chain pairing preferences in the parallel coiled-coil dimer motif: insight on ion pairing between core and flanking sites. *J. Am. Chem. Soc.* 132, 7586–7588.
- (39) Steinkruger, J. D., Bartlett, G. J., Woolfson, D. N., and Gellman, S. H. (2012) Strong contributions from vertical triads to helix-partner preferences in parallel coiled coils. *J. Am. Chem. Soc.* 134, 15652–15655.

Supporting Information

Software - The following tools are listed in the order that have been used to derive the heterospecific CCs listed in the manuscript.

i) ***Generate Library Sequences***

<http://people.bath.ac.uk/jm2219/biology/create-library.htm> generates a list of all given sequences within a peptide library. Data can be entered at the peptide (where a '?' defines each position to be scrambled with specific options for each '?' entered by the user) or DNA level (where libraries are built using degenerate codons (e.g. the codon WSG = options of S,T,W,R)).

ii) ***bcIPA Interactome Screen:***

<http://people.bath.ac.uk/jm2219/biology/bcipa-interactome.htm> screens all interactions between a defined set of sequences (generated from i) and outputs a predicted T_m value for each interaction. This includes an option to provide a colour-coded heat-map of the resulting interactome. In the case of our 1536 member library this represented 1,180,416 hypothetical PPIs within the interactome.

iii) ***Find Pairs / Find Quadruples / Find Octuples:***

<http://people.bath.ac.uk/jm2219/biology/find-pairs.php> works using the ***bcIPA Interactome Screen*** engine. It allows users to input a library of sequences and then screen the resulting interactome to identify two (i.e. four peptides - pairs), four (i.e. eight peptides - quadruples), or even higher numbers of sets of leucine zippers that are specific within each other's presence. Users can also input desired parameters for their specific pairs (maximum homodimer T_m , minimum desired T_m , maximum undesired T_m , minimum delta T_m) depending on their requirements. The software is incremental, meaning that options for 'find quadruples' will appear within the software once 'find pairs' has identified sequences that meet the given requirements. Throughout, the user is able to control stringency by inputting the necessary parameters. These include the desired ΔT_m , maximum undesired T_m and maximum desired T_m . The output will then list all T_m values within the resultant eight-peptide interactome and generate a heatmap (Figure 2). Required parameters for screening the complete 1536 peptide interactome screen are listed in the methods section and are designed to keep homodimeric and off-target T_m values low, and desired T_m values high. Having screened all 1,180,416 hypothetical interactions we were able to identify pairs, and later quadruples of coiled coils that met the specificity criteria. Once sequences were identified within ***Find pairs***, the ***Find***

quadruples options were available to identify sets of pairs that retained specificity when combined into quadruples. Finally, once sequences were identified within **Find Quadruples**, the **Find Octuples** options were available to identify sets of Quadruples that retained specificity when combined into Octuples. Again, the user is able to input a maximum tolerated off-target T_m value as well as a minimum difference in T_m between desired and non-desired pairs.

Setting T_m cut-off points. Having generated the peptide library, a second program ('*Find pairs*') searches within the imported interactome by integrating with bCIPA to predict of PPI T_m values and identify pairs of coiled coils (i.e. four peptides) that are expected to be specific within each other's presence. Once identified, the program searches through pairs of coiled coils to identify four coiled coils (i.e. eight peptides - Quadruples). This is followed by combining Quadruples to identify Octuples – sets of 16 peptides predicted to form 8 heterospecific coiled coils. The program therefore works iteratively by identifying which interactions between a set of sequences do and do not take place given a set of cut-off parameters input by the user. The program assigns a T_m to every hypothetical pair and creates an associated heatmap for the resulting interactome. The stringency for desired pairs can be changed according to required stability, as well as the T_m cut-off for off-target interactions (i.e. homodimers and heterodimers). Given these sets of customised constraints, first Quadruples and next Octuples can be identified. In our case the maximum acceptable predicted T_m for homodimers was set to 10 °C, the maximum T_m for desired heterodimers was set to 70 °C, the maximum T_m for undesired heterodimers was set as 20 °C, with the minimum ΔT_m (desired – off-target) set as 50 °C. After enabling the antiparallel coiled coil removal feature, this resulted in 42 separate groups of predicted non-interacting pairs of coiled coils. The highest predicted T_m for the desired coiled coils was 73 °C and the highest predicted T_m for undesired coiled coils was 18 °C. The software additionally allows the user to specify other desired characteristics – in this case that two Asn residues and two Ile residues should be found within every peptide to direct specificity¹⁻³. Having identified two coiled coils predicted to be of both high affinity and specificity, the program can be further instructed to identify four coiled coils (Quadruples). In this case, minimum ΔT_m of 40 °C, and a maximum undesired coiled coil T_m of 30 °C was specified within the software. This resulted in retaining 72 groups of non-interacting Quadruples of coiled coils with the highest desired T_m of 73 °C and a lowest undesired coiled coil T_m of 28 °C. Finally the parameters can be set to identify eight coiled coils (Octuples) using a minimum ΔT_m of 21 °C, and a maximum undesired coiled coil T_m of 52 °C. This resulted in 36 groups of non-interacting Octuples of coiled coils, with the highest desired T_m of 73 °C and a T_m of 52 °C for the lowest undesired coiled coil.

Python-based Software. The PHP-based software for ‘*Find X with Coils*’ was used to increase the speed of the search and to allow the search to become more expansive when using either bCIPA or qCIPA to identify heterospecific sets of coiled coils. The software does not currently run via html but is freely available from j.mason@bath.ac.uk upon request.

Peptides Used in Quadruple Training Set (see Crooks et al 2016 ⁴):

Quad01: ASENAALEAKNAALKYKIAALKAEIAALEGAP
 Quad02: ASKNAALKAENAALEYEIAALEAKIAALKGAP
 Quad03: ASEIAALEAEIAALEYENAALEAENAALEGAP
 Quad04: ASKIAALKAKIAALKYKNAALKAKNAALKGAP
 Quad05: ASKNAALKAETIAALEYKIAALKAEENAALEGAP
 Quad06: ASENAALEAKIAALKYEIAALEAKNAALKGAP
 Quad07: ASKIAALKAKNAALKYENAALEAEIAALEGAP
 Quad08: ASEIAALEAENAALEYKNAALKAKIAALKGAP

Peptides Used in Octuples Test Set:

Oct01: ASKNAALKAENAALEYEIAALEAEIAALEGAP
 Oct02: ASENAALEAKNAALKYKIAALKAKIAALKGAP
 Oct03: ASEIAALEAKIAALKYENAALEAENAALEGAP
 Oct04: ASKIAALKAEIAALEYKNAALKAKNAALKGAP
 Oct05: ASENAALEAEIAALEYKNAALKAEIAALEGAP
 Oct06: ASKNAALKAKIAALKYENAALEAKIAALKGAP
 Oct07: ASKIAALKAKNAALKYKIAALKAEENAALEGAP
 Oct08: ASEIAALEAENAALEYEIAALEAKNAALKGAP
 Oct09: ASENAALEAKNAALKYEIAALEAEIAALEGAP
 Oct010: ASKNAALKAENAALEYKIAALKAKIAALKGAP
 Oct011: ASKIAALKAEIAALEYENAALEAENAALEGAP
 Oct012: ASEIAALEAKIAALKYKNAALKAKNAALKGAP
 Oct013: ASKNAALKAKIAALKYKNAALKAEIAALEGAP
 Oct014: ASENAALEAEIAALEYENAALEAKIAALKGAP
 Oct015: ASEIAALEAENAALEYKIAALKAEENAALEGAP
 Oct016: ASKIAALKAKNAALKYEIAALEAKNAALKGAP

References

- [1] Lumb, K. J., and Kim, P. S. (1995) A Buried Polar Interaction Imparts Structural Uniqueness in a Designed Heterodimeric Coiled-Coil, *Biochemistry* 34, 8642-8648.
- [2] Oakley, M. G., and Kim, P. S. (1998) A buried polar interaction can direct the relative orientation of helices in a coiled coil, *Biochemistry* 37, 12603-12610.
- [3] Gonzalez, L., Jr., Woolfson, D. N., and Alber, T. (1996) Buried polar residues and structural specificity in the GCN4 leucine zipper, *Nat. Struct. Biol.* 3, 1011-1018.
- [4] Crooks, R. O., Baxter, D., Panek, A. S., Lubben, A. T., and Mason, J. M. (2016) Deriving Heterospecific Self-Assembling Protein-Protein Interactions Using a Computational Interactome Screen, *J. Mol. Biol.* 428, 385-398.

Type of ES Pairing	Core Type									
	NN NN II II	NN II NN II	II NN II II	NN II II NN	NN II II NN	II NN NN II	II II NN NN	II II NN NN	II II NN NN	II II NN NN
EK EK EK EK									Q3-4	75
EK EK KE EK		O5-6	71	O15-16	52					
EK KE EK EK	O9-10	69							O3-4	68
KE EK EK EK	O1-2	72							O11-12	73
KE KE KE EK			O13-14	75	O7-8	68				
KE EK KE EK							Q5-6	58		
KE KE EK EK									Q7-8	64
EK KE KE EK	Q1-2	63								

Table S1: Shown are all desired pairs (including parallel dimers from Crooks *et al* 2016⁴) for comparison. These pairs have 2II 2NN core arrangements and all attractive electrostatics arrangements. By comparing data in the same row or column we are able to make some limited inferences on the effect of different core or electrostatic arrangements, respectively. For example O5-6 > O15-16 ($\Delta T_m = 19$) and O13-14 > O7-8 ($\Delta T_m = 7$) suggesting that NN II NN II arrangement leads to increased stability than II NN II NN. In contrast O9-10 \approx O3-4 and O1-2 \approx O11-12, suggesting that NN NN II II and II II NN NN core arrangements are energetically equal. Looking at electrostatics, O13-14 > O5-6 ($\Delta T_m = 4$) and 7-8 > O15-16 ($\Delta T_m = 16$) which together suggest that the KE KE KE EK arrangement is more stable than EK EK KE EK. Similarly O1-2 > O9-10 ($\Delta T_m = 3$) and O11-12 > O3-4 ($\Delta T_m = 5$), both suggesting that KE EK EK EK is more stable than EK KE EK EK. Taking this further both O1-2 and O9-10 are favoured over Q1-2 ($\Delta T_m = 9$ and 6) suggesting that EK KE KE EK is less stable than either alternative arrangement. Collectively this suggests that for desired pairs blocks of same charge on either *e* or *g* residues of each peptide lead to increased stability over that purely based on the sum of the core and electrostatic components (vinson refs). We speculate that these 'charge blocks' increase stability for these desired pairs by promoting inter-molecular attraction.

Type of ES Repulsion	Core Type		Average Tm
	NI IN NI IN	NI NI IN IN	
+ + . .	O6-7 O13-16	19 16	18 ± 2
. + + .		Q1-4 28	28 ± 0
. . + +		O2-4 21 O10-12 26	24 ± 3
+ . . +		Q2-4 31	31 ± 0
- - . .	O5-8 O14-15	26 20	23 ± 3
. - - .		Q2-3 20	20 ± 0
. . - -		O1-3 3 O9-11 5	4 ± 1
- . . -		Q1-3 11	11 ± 0
+ - . .		O1-4 20 O10-11 31	26 ± 6
- + . .		O2-3 27 O9-12 42	35 ± 8
. . + -	O5-7 O13-15	35 51	43 ± 8
. . - +	O6-8 O14-16	48 34	41 ± 7

Table S3: Shown are all intra-pair off-targets (including parallel dimers from Crooks *et al* 2016⁴) for comparison. These pairs contain 4NI core mismatches and 4 attractive / 4 repulsive electrostatic arrangements. No comparisons can be made between the two core types. A . signifies either an EK or KE electrostatic attraction. By comparing electrostatics with a normalised core we can see that when alternative polarity electrostatic charge repulsions flank each other this again leads to increased stability. Similarly when same polarity charge repulsions flank each other (i.e. ++ or --) we observe a decrease in stability. As previously, we speculate that these promote intra-molecular attraction and cause decreased $e\text{-}g^{+/+}$ inter-molecular repulsion. These effects increase stability for these intra-pair off-targets.

Type of ES Repulsion	Core Type														Average Tm		
	NN	NI	IN	II	NI	NN	II	IN	II	NN	NI	II	IN	NI		NN	
+ + . . + . + . + . . + . + + . . + . + . . + +														Q4-7	16	16 ± 0	
	O10-13	26						O4-13	28						O4-7	16	25 ± 3
	O6-10	27	O10-16	28				O4-6	38						O4-16	15	27 ± 5
	O2-13	13	O2-7	15				O12-13	20						O7-12	15	16 ± 1
	O2-6	30	O2-16	18						O6-12	30				O12-16	21	25 ± 3
														Q4-8	33	33 ± 0	
								O3-14	30					Q3-8	20	25 ± 5	
- . . . - . - . - . . - . - - . . - . - . . - -	O9-14	18														20 ± 2	
	O5-9	11	O9-15	20				O3-5	30						O3-15	6	17 ± 5
	O1-14	23	O1-8	-3				O11-14	17						O8-11	5	11 ± 6
	O1-5	21	O1-15	8						O5-11	15				O11-15	13	14 ± 3
														Q3-7	18	18 ± 0	
	O1-6	23	O1-16	14				O11-13	26				O6-11	23	O11-16	4	21 ± 2
	O1-13	43	O1-7	34											O7-11	21	33 ± 6
. + - . . + . - . + . - . + . - . . + -	O6-9	40	O9-16	18				O3-6	33						O3-16	21	28 ± 5
	O9-13	32	O7-9	17											O3-7	18	23 ± 4
								O3-13	22								22 ± 0
	O2-5	30	O2-15	42						O5-12	29				O12-15	26	32 ± 3
	O2-14	54	O2-8	51				O12-14	51						O8-12	48	51 ± 1
	O5-10	48	O10-15	26				O4-5	36						O4-15	40	38 ± 5
	O10-14	31	O8-10	29				O2-8	47				O4-14	36	O4-8	22	33 ± 5

Table S4: Shown are all Inter-Pair off-targets type I (including parallel dimers from Crooks *et al* 2016⁴) for comparison. These contain III INN 2NI core arrangements and 4 attractive / 4 repulsive electrostatic arrangements. The most stable of these off-targets by some margin are those with a +.- configuration, suggesting that these help to offset the helix macrodipole. This is particularly stabilising when a negative N-terminus or + C-terminus is flanked by attractive (charge neutral) electrostatic pairs. Another general observation is that the NN NI IN II, NI NN II IN and IN II NN NI core configurations are generally more stable than the II IN NI NN configuration. Again when same polarity charge repulsions flank each other at e or g positions (i.e. ++ or --) we observe a decrease in stability relative to opposing charges (+- or -+). As previously, we speculate that these promote intra-molecular attraction and cause decreased $e\text{-}g^{+/}$ inter-molecular repulsion. Therefore in a similar but opposite manner to the desired dimers, any ‘charge blocks’ decrease the stability for off-targets pairs by promoting inter-molecular repulsion. Conversely, alternating the polarity of the electrostatic residues at the *e* and *g* positions leads and increases in dimer stability because it causes less inter-molecular repulsion and promotes intra-molecular attraction.

Type of ES Repulsion	Core Type					Average T _m
	IN NI	IN NI	NI NI	IN IN	IN NI	
EK EK EK EK						
EK EK EK KE	08-13	42				42 ± 0
EK EK KE EK					05-16 21	21 ± 0
EK KE EK EK				03-10 10		10 ± 0
KE EK EK EK			01-12 36			36 ± 0
KE KE KE EK	07-14	56				56 ± 0
KE KE EK KE			06-15 31			31 ± 0
KE EK KE KE				04-9 28		28 ± 0
EK KE KE KE			02-11 38			38 ± 0

Table S6: Shown are all Inter-Pair off-targets type III (including parallel dimers from Crooks *et al* 2016⁴) for comparison. These contain 4NI core arrangements and all attractive electrostatic components). These have the same electrostatic arrangements as the desired coiled coils and once again we observe that same charge residues at consecutive e or g positions results in increased stability.

Type of ES Repulsion	Core Type						Average Tm
	IN NI	IN NI	NI NI	IN IN	NI NI	NI NI	
+ - - -			01-11	1			1 ± 0
- + - -					03-9	18	18 ± 0
- - + -						05-15 36	36 ± 0
- - - +	08-14	17					17 ± 0
- - - -							
- + + +			02-12	17			17 ± 0
+ - + +					04-10 32		32 ± 0
+ + - +						06-16 25	25 ± 0
+ + + -	07-13	18					18 ± 0
+ + + +							

Table S7: Shown are all Inter-Pair off-targets type IV (including parallel dimers from Crooks *et al* 2016⁴) for comparison. These contain 4NI core arrangements and all repulsive electrostatic components. Once again intra-molecular charge attraction between $e-e^{+I}$ or $g-g^{+I}$ can add additional stability to these off-targets while off-setting the intended inter-molecular repulsion.

	Peptide Mixture (T _m)	Elution Time 50 μ M	Elution Time 10 μ M
Individual Peptides	1=1 (-2)	19.91	N/A
	2=2 (34)	20.78	N/A
	3=3 (13)	20.09	N/A
	4=4 (21)	20.12	N/A
	5=5 (17)	20.06	N/A
	6=6 (23)	20.35	N/A
	7=7 (-1)	19.99	N/A
	8=8 (2)	19.97	N/A
	9=9 (6)	19.99	N/A
	10=10 (29)	20.18	N/A
	11=11 (4)	19.71	N/A
	12=12 (15)	20.26	N/A
	13=13 (21)	20.60	N/A
	14=14 (24)	20.12	N/A
	15=15 (12)	20.00	N/A
	16=16 (0)	20.18	N/A
Desired	1=2 (72)	-1.40 \pm 0.43 (18.95)	-1.18 \pm 0.43 (19.17)
	3=4 (68)	-1.08 \pm 0.02 (19.03)	-1.00 \pm 0.02 (19.10)
	5=6 (71)	-1.23 \pm 0.15 (18.98)	-0.88 \pm 0.15 (19.33)
	7=8 (68)	-1.11 \pm 0.01 (18.87)	-0.76 \pm 0.01 (19.22)
	9=10 (69)	-1.16 \pm 0.10 (18.93)	-0.92 \pm 0.10 (19.17)
	11=12 (73)	-1.00 \pm 0.28 (18.99)	-0.94 \pm 0.28 (19.05)
	13=14 (75)	-1.51 \pm 0.24 (18.85)	-1.19 \pm 0.24 (19.17)
	15=16 (52)	-1.08 \pm 0.10 (19.01)	-0.04 \pm 0.10 (20.05)
Off-Target	1=7 (34)	-0.07 \pm 0.04 (19.88)	N/A
	1=12 (36)	-0.20 \pm 0.18 (19.89)	N/A
	2=5 (30)	-0.08 \pm 0.36 (20.50)	N/A
	2=8 (51)	-0.52 \pm 0.41 (19.86)	-0.21 \pm 0.41 (20.17)
	2=9 (56)	-0.81 \pm 0.40 (19.58)	-0.45 \pm 0.40 (19.93)
	2=10 (46)	0.00 \pm 0.30 (20.48)	N/A
	2=14 (54)	-0.83 \pm 0.33 (19.62)	-0.28 \pm 0.33 (20.17)
	4=10 (32)	-0.03 \pm 0.03 (20.18)	N/A
	12=14 (51)	-0.52 \pm 0.07 (19.67)	0.04 \pm 0.07 (23.23)
	3=12 (60)	-0.21 \pm 0.09 (19.97)	N/A
	4=11 (50)	0.19 \pm 0.21 (20.10)	N/A
	6=14 (55)	-0.55 \pm 0.11 (19.67)	-0.55 \pm 0.11 (19.69)
	7=14 (56)	-1.10 \pm 0.07 (19.00)	-0.13 \pm 0.07 (19.93)
	13=15 (51)	-0.39 \pm 0.30 (19.91)	N/A

Table s8: Size exclusion chromatography experiments. Shown are elution times for interacting and non-interacting peptides. For heteromeric samples the difference in elution time is shown. For example, homomeric samples eluted at 20 minutes. In contrast, a peak at 19 minutes indicated that component peptides have dimerised. This was found to be the case in intended samples 1-2, 3-4, 5-6, 7-8, 9-10, 11-12, 13-14, and 15-16 at 50 μ M Pt. Previously characterised controls eluted at 20 min for a monomeric sample (a 32mer Fos sample) and 18.5 min for a 37mer cJun-FosW heterodimer. The table summarises elution times for all homomeric samples as well as all 8 intended dimeric samples and a range of off-target heteromeric samples. Data are shown at 50 μ M and 10 μ M Pt. Of the intended dimers only 15-16 is no longer dimeric at 10 μ M Pt. However, at this concentration all other off-target samples (including all off-target samples with a T_m > 50 $^{\circ}$ C) were found to be monomeric. These experiments provide strong evidence for heterospecificity of the other 7 designed coiled coils.

		bCIPA correlation	qCIPA correlation
Quadruples training set	Overall r^2	0.60	0.69
	Peptides 1-4	0.69	0.85
	Peptides 5-8	0.93	0.93
	Inter	0.00	0.00
Octuples validation set			
	Overall r^2	0.31	0.36
	Peptides 1-4	0.71	0.89
	Peptides 5-8	0.82	0.69
	Peptides 9-12	0.71	0.84
	Peptides 13-16	0.70	0.59
	Inter 1	0.05	0.04
	Inter 2	0.12	0.26
	Inter 3	0.00	0.00
	Inter 4	0.13	0.10
	Inter 5	0.21	0.18
	Inter 6	0.06	0.05

Table s9: Table of correlation coefficients between observed T_m values and those predicted by bCIPA and qCIPA. For qCIPA the quadruple peptides were used as a training set and the subsequent fit applied to the octuple validation set. Overall correlations are shown as well as intra-pair and inter-pair values.